

# THE INFINITUDE OF THE PRIMES

KEITH CONRAD

## 1. INTRODUCTION

The sequence of prime numbers

2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59,  $\dots$ , 1873, 1877, 1879, 1889, 1901,  $\dots$

never ends. This fact has many different proofs. We'll discuss two of them, due to Euclid and to Euler, and some of their consequences.

## 2. EUCLID'S PROOF

The standard proof of the infinitude of the primes is attributed to Euclid and uses the fact that all integers greater than 1 have a prime factor.

**Lemma 2.1.** *Every integer greater than 1 has a prime factor.*

*Proof.* We argue by (strong) induction that each integer  $n > 1$  has a prime factor. For the base case  $n = 2$ , 2 is prime and is a factor of itself.

Now assume  $n > 2$  all integers greater than 1 and less than  $n$  have a prime factor. To show  $n$  has a prime factor, we take cases.

Case 1:  $n$  is prime.

Since  $n$  is a factor of itself,  $n$  has a prime factor when  $n$  is prime.

Case 2:  $n$  is not prime.

Since  $n$  is not prime, it has a factorization  $n = ab$  where  $1 < a, b < n$ . Then by the strong inductive hypothesis,  $a$  has a prime factor, say  $p$ . Since  $p \mid a$  and  $a \mid n$ , also  $p \mid n$  and thus  $n$  has prime factor  $p$ . □

**Theorem 2.2.** *There are infinitely many primes.*

*Proof.* (Euclid) To show there are infinitely many primes, we'll show that every finite list of primes is missing a prime number, so the list of all primes can't be finite.

To begin, there are prime numbers such as 2. Suppose  $p_1, \dots, p_r$  is a finite list of prime numbers. We want to show this is not the full list of the primes. Consider the number

$$N = p_1 \cdots p_r + 1.$$

Since  $N > 1$ , it has a prime factor  $p$  by Lemma 2.1. The prime  $p$  can't be any of  $p_1, \dots, p_r$  since  $N$  is not divisible by any of those primes. Therefore  $p$  is a prime not on our list, so the set of primes can't be finite. □

Some people misunderstand this proof to be saying that if  $p_1, \dots, p_r$  are prime then  $p_1 \cdots p_r + 1$  is prime. That is not generally true. It starts out looking correct:  $2 + 1$  is

prime,  $2 \cdot 3 + 1 = 7$  is prime,  $2 \cdot 3 \cdot 5 + 1 = 31$  is prime,  $2 \cdot 3 \cdot 5 \cdot 7 + 1 = 211$  is prime, and  $2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 + 1 = 2311$  is prime, but

$$2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 + 1 = 30031 = 59 \cdot 509$$

is *not* prime. (More simply,  $2 \cdot 7 + 1 = 15$  and  $3 \cdot 5 + 1 = 16$  are not prime.) Euclid's proof tells us that we can always find a prime outside of a finite list of primes  $p_1, \dots, p_r$  by using a prime factor of  $p_1 \cdots p_r + 1$ , not by using  $p_1 \cdots p_r + 1$  itself.

Let's try to use Euclid's proof as a method of finding new primes, by taking the *smallest* prime factor of  $p_1 \cdots p_r + 1$  at each step as a new prime to add to our list:

- 2 is prime and  $2 + 1 = 3$  is prime,
- $2 \cdot 3 + 1 = 7$  is prime,
- $2 \cdot 3 \cdot 7 + 1 = 43$  is prime,
- $2 \cdot 3 \cdot 7 \cdot 43 + 1 = 1807 = 13 \cdot 139$ ,
- $2 \cdot 3 \cdot 7 \cdot 43 \cdot 13 + 1 = 23479 = 53 \cdot 443$ ,
- $2 \cdot 3 \cdot 7 \cdot 43 \cdot 13 \cdot 53 + 1 = 1244335 = 5 \cdot 248867$ .

So far this list of primes in the order they appear is 2, 3, 7, 43, 13, 53, 5. This way of creating new primes was introduced by Mullin [4] in 1963 and is called the Euclid–Mullen sequence. Further terms in this sequence are on the OEIS webpage <https://oeis.org/A000945>: 11 is the 12th term, 17 is the 13th term, 19 is the 36th term, and 23 is the 25th term. Mullin asked if every prime number actually appears as some term in this sequence; this is an unsolved problem.

Mullin also considered a similar algorithm to generate new primes by using the *largest* prime factor of  $p_1 \cdots p_r + 1$  at each step. This “second Euclid–Mullen sequence” starts off as 2, 3, 7, 43, 139, 50207, 340999. If this list is continued, some primes definitely never show up in it, such as all the primes from 5 to 47 [2]. In fact, infinitely many primes are missing from this sequence [1], [5].

The number of primes up to  $x$  is written as  $\pi(x)$ . For example,  $\pi(10) = |\{2, 3, 5, 7\}| = 4$  and  $\pi(12.7) = |\{2, 3, 5, 7, 11\}| = 5$ . Because there are infinitely many primes,  $\pi(x) \rightarrow \infty$  as  $x \rightarrow \infty$ . Euclid's proof gives us the following crude lower bound on the number of primes up to  $x$ .

**Corollary 2.3.** *For  $x \geq 2$ ,  $\pi(x) > \log_2(\log_2 x) - 1$ .*

*Proof.* If  $p_1, \dots, p_n$  are the first  $n$  primes, then Euclid's proof tells us that we can find a new prime using a prime factor of  $p_1 \cdots p_n + 1$ . Therefore  $p_{n+1} \leq p_1 \cdots p_n + 1$ . Using this inequality, we can show by induction that  $p_n < 2^{2^n}$  for all  $n \geq 1$ : it's true when  $n = 1$  since  $2 < 2^{2^1}$ , and if that upper bound on primes is true for  $p_1, \dots, p_k$  then

$$p_{k+1} \leq p_1 p_2 \cdots p_k + 1 < 2 \cdot 2^2 \cdots 2^{2^k} + 1 = 2^{1+2+\cdots+2^k} + 1 = 2^{2^{k+1}-1} + 1,$$

and that last number is less than  $2^{2^{k+1}}$  since  $\frac{1}{2}x + 1 < x$  for  $x > 2$  (take  $x = 2^{k+1}$ ).

That  $p_n < 2^{2^n}$  for all  $n \geq 1$  means the number of primes below  $2^{2^n}$  is at least  $n$ , so  $\pi(2^{2^n}) \geq n$  for  $n \geq 1$ . For  $x \geq 2$ , choose the positive integer  $n$  such that  $2^n \leq \log_2 x < 2^{n+1}$ . Then  $x \geq 2^{2^n}$ , so  $\pi(x) \geq \pi(2^{2^n}) \geq n > \log_2(\log_2 x) - 1$ .  $\square$

This lower bound on  $\pi(x)$  is extremely weak because the true order of magnitude of  $\pi(x)$  is  $x/\log x$ : the ratio  $\pi(x)/(x/\log x)$  tends to 1 as  $x \rightarrow \infty$ . That is called the Prime Number Theorem.

## 3. EULER'S PROOF

In 1737, Euler [3, Theorem 7] found a proof of the infinitude of the primes that explains it by the divergence of the harmonic series. This unexpected link between a property of prime numbers and calculus (infinite series) could be considered the start of the subject of *analytic number theory*, which studies properties of  $\mathbf{Z}$  using the tools of real and complex analysis. While Euclid's proof used the fact that each integer greater than 1 has a prime factor, Euler's proof will rely on unique factorization in  $\mathbf{Z}^+$ .

**Theorem 3.1.** *There are infinitely many primes.*

*Proof.* (Euler) For a prime  $p$ , the ratio  $1/(1 - 1/p)$  can be expanded into a geometric series:

$$\frac{1}{1 - 1/p} = 1 + \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \frac{1}{p^4} + \cdots$$

We will now multiply together these ratios for different primes. To see what can be obtained, let's look at the product of these terms for the primes 2, 3, and 5:  $\frac{1}{1-1/2} \frac{1}{1-1/3} \frac{1}{1-1/5}$  equals

$$\left(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots\right) \left(1 + \frac{1}{3} + \frac{1}{9} + \frac{1}{27} + \cdots\right) \left(1 + \frac{1}{5} + \frac{1}{25} + \frac{1}{125} + \cdots\right),$$

and if we multiply together one term from each series we will get unit fractions  $1/n$  where  $n$  is a product of powers of 2, 3, and 5:

$$(3.1) \quad 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{8} + \frac{1}{9} + \frac{1}{10} + \frac{1}{12} + \frac{1}{15} + \frac{1}{16} + \frac{1}{18} + \frac{1}{20} + \frac{1}{24} + \cdots$$

This looks like the harmonic series, but it is missing terms  $1/n$  where  $n$  has a prime factor greater than 5, such as terms  $1/7$ ,  $1/11$ , and  $1/14$ . If we multiply by additional factors  $1/(1 - 1/p)$  for more primes  $p$  we'll introduce into (3.1) new terms  $1/n$  where  $n$  has prime factors involving the new primes and the ones we already used (2, 3, and 5). Because of unique factorization in  $\mathbf{Z}^+$ , each possible term  $1/n$  that might appear will do so just one time. Therefore if we multiply together  $1/(1 - 1/p)$  for *all* the primes, (3.1) turns into the sum of  $1/n$  over all positive integers  $n$ :

$$(3.2) \quad \prod_p \frac{1}{1 - 1/p} = \sum_{n \geq 1} \frac{1}{n},$$

where the product on the left (denoted  $\Pi$ , like  $\Sigma$  for sum) runs over all prime numbers.

Since the harmonic series diverges, (3.2) tells us that the left side can *not* be a product of finitely many terms. Therefore there are infinitely many terms in the product, so there are infinitely many primes.  $\square$

By taking logarithms of a product over primes, Euler showed something new about primes when we sum their reciprocals:

**Corollary 3.2.** *The infinite series  $\sum_p 1/p$ , where  $p$  runs over the primes, diverges.*

*Proof.* For  $N \geq 2$ , multiply out

$$\prod_{p \leq N} \frac{1}{1 - 1/p} = \prod_{p \leq N} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \frac{1}{p^4} + \cdots\right),$$

where the product is over primes  $p$  up to  $N$ , by multiplying together one term from each of the geometric series on the right. We obtain unit fractions  $1/n$  that include every integer

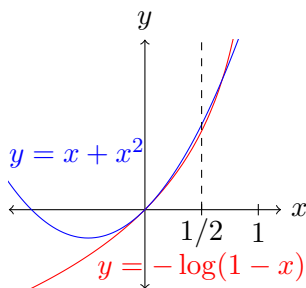
$n \leq N$ , since the prime factors of such  $n$  are all at most  $N$ . We'll also get unit fractions  $1/n$  for other  $n$ . Therefore

$$(3.3) \quad \sum_{n \leq N} \frac{1}{n} < \prod_{p \leq N} \frac{1}{1 - 1/p}$$

As  $N \rightarrow \infty$  the left side of (3.3) tends to  $\infty$  as  $N$  does, so the right side of (3.3) tends to  $\infty$  too. That isn't really telling us anything new, since (3.2) already says  $\prod_p \frac{1}{1-1/p} = \infty$ . We can get something new from (3.3) by taking logarithms of both sides before letting  $N$  tend to  $\infty$ : logarithms turn products into sums and  $\log x$  is increasing, so (3.3) implies

$$(3.4) \quad \log \left( \sum_{n \leq N} \frac{1}{n} \right) \leq \sum_{p \leq N} \log \frac{1}{1 - 1/p}$$

On the right,  $\log(1/(1 - 1/p)) = -\log(1 - 1/p)$ . Recall for  $|x| < 1$  that the power series for  $-\log(1 - x)$  is  $\sum_{n \geq 1} x^n/n = x + x^2/2 + x^3/3 + \dots$ , so for small  $x$  we have  $-\log(1 - x) \approx x + x^2/2$ . We need a definite inequality here, not just " $\approx$ ", and for that increase  $x^2/2$  up to  $x^2$ : for  $0 < x \leq 1/2$ , verify by calculus that  $-\log(1 - x) < x + x^2$  (see the graph below).



Using  $x = 1/p$  for prime  $p$ , so  $x \leq 1/2$ , we have  $-\log(1 - 1/p) < 1/p + 1/p^2$ . Thus

$$\sum_{p \leq N} -\log(1 - 1/p) < \sum_{p \leq N} \left( \frac{1}{p} + \frac{1}{p^2} \right).$$

Combining this with (3.4),

$$(3.5) \quad \log \left( \sum_{n \leq N} \frac{1}{n} \right) < \sum_{p \leq N} \frac{1}{p} + \sum_{p \leq N} \frac{1}{p^2}.$$

Now let  $N \rightarrow \infty$ : the left side tends to  $\infty$  since  $\sum_{n \leq N} 1/n \rightarrow \infty$  and on the right side  $\sum_{p \leq N} 1/p^2$  has a finite limit as  $N \rightarrow \infty$  since  $\sum_{n \geq 1} 1/n^2$  converges. Therefore (3.5) implies  $\sum_{p \leq N} 1/p \rightarrow \infty$  as  $N \rightarrow \infty$ , so  $\sum_p 1/p = \infty$ .  $\square$

Since  $\sum_{n \geq 1} 1/n = \infty$  and  $\sum_p 1/p = \infty$  while  $\sum_{n \geq 1} 1/n^2 < \infty$ , the primes are "more dense" among the positive integers than the squares.

Although  $\sum_p 1/p = \infty$ , the divergence is *extremely* slow. For example,  $\sum_{p < 10^8} 1/p \approx 3.174975$ . The partial sums  $\sum_{p \leq N} 1/p$  diverge at the rate  $\log(\log N)$ , which Euler knew: the last line of his paper where he proved infinitude of the primes [3] says

$$\frac{1}{2} + \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{11} + \text{etc.} = l.l \infty,$$

where  $l$  was Euler's notation for natural logarithms.

In a MathOverflow post about the importance of rigor<sup>1</sup>, the top answer by Richard Borcherds says "I once got a letter from someone who had overwhelming numerical evidence that the sum of the reciprocals of primes is slightly bigger than 3 (he may have conjectured the limit was  $\pi$ ). The sum is in fact infinite, but diverges so slowly (like  $\log \log n$ ) that one gets no hint of this by computation." We saw  $\sum_{p < 10^8} 1/p$  is slightly above 3. Since the solution to  $\log(\log x) = 4$  is a 24-digit number, a calculator or computer will never give convincing evidence of the divergence of  $\sum_p 1/p$ .

Corollary 3.2 was very important historically in number theory: it suggests that a way to prove a set  $S$  of primes is infinite is to show the sum  $\sum_{p \in S} 1/p$  is  $\infty$ . In fact, 100 years after Euler's work Dirichlet used this method in 1837 to prove certain arithmetic progressions contain infinitely many primes.

**Theorem 3.3** (Dirichlet). *If  $a$  and  $m$  are relatively prime positive integers, there are infinitely many primes in the arithmetic progression  $\{a + mn : n \geq 0\}$ . Equivalently, there are infinitely many primes  $p$  such that  $p \equiv a \pmod{m}$ .*

For example, the set of primes  $p$  with units digit 7, namely  $p \equiv 7 \pmod{10}$ , is infinite. It starts out as

$$17, 37, 47, 57, 67, 97, 107, 127, 137, 157, 167, 197, 227, 257, 277, \dots$$

The basic idea in Dirichlet's proof is to show the series  $\sum_{p \equiv a \pmod{m}} 1/p$  over primes  $p$  is infinite when  $(a, m) = 1$ , so the set of primes  $p \equiv a \pmod{m}$  is infinite. Some special cases of Dirichlet's theorem (such as infinitude of primes  $p$  where  $p \equiv 3 \pmod{4}$ ) can be proved by elementary techniques, but I am unaware of any elementary method that shows the list of primes  $p \equiv 7 \pmod{10}$  is infinite.

A famous unsolved problem in number theory is the infinitude of *twin primes*, which are prime pairs that differ by 2, such as 3 and 5 or 29 and 31. If  $T$  is the set of primes  $p$  such that  $p + 2$  is also prime, might  $\sum_{p \in T} 1/p$  be infinite? That would imply there are infinitely many twin primes. Unfortunately,  $\sum_{p \in T} 1/p$  is known to converge (which doesn't contradict the possibility that  $T$  is infinite: infinite series can converge). The convergence of  $\sum_{p \in T} 1/p$  was proved by Brun in 1919, so  $\sum_{p \in T} 1/p$  is called Brun's constant. Estimating this number is a difficult problem, and work on this in 1994 led to the discovery of a [bug in an Intel Pentium chip](#).

## REFERENCES

- [1] A. Booker, "On Mullin's second sequence of primes," *Integers* **12** (2012), 1167–1177.
- [2] C. D. Cox and A. J. van der Poorten, "On a sequence of prime numbers," *J. Austral. Math. Soc.* **8** (1968), 571–574.
- [3] L. Euler, "Variae observationes circa series infinitas," *Commentarii academiae scientiarum Petropolitanae* **9** (1744), 160–188. Online at <http://eulerarchive.maa.org/pages/E072.html>.
- [4] A. A. Mullin, "Recursive function theory (A modern look at a Euclidean idea)," *Bull. Amer. Math. Soc.* **69** (1963), 737.
- [5] P. Pollack and E. Treviño, "The primes that Euclid forgot," *Amer. Math. Monthly* **121** (2014), 433–437.
- [6] P. Ribenboim, *The Book of Prime Number Records*, Springer-Verlag, New York, 1988.

<sup>1</sup>See <https://mathoverflow.net/questions/37610>.