Primes

1 Infinity of Primes – Five Proofs

In this section we collect various proofs of the infinity of primes. Most of the material is from PTB. The proofs are more or less in the chronological order of discovery.

¶1. Euclid: Let $p_1, \ldots, p_k$ be the finite set of primes. Consider the number $N := p_1 \ldots p_k + 1$. None of the primes $p_1, \ldots, p_k$ divide it. So either N is a prime, or it is divisible by a prime number distinct from $p_1, \ldots, p_k$.

Remakr: Kummer simplified the argument slightly as follows: Let $N := p_1 \ldots p_k$; then N and $N + 1$ share a prime p; thus p divides their difference, which is one, giving us a contradiction.

¶2. Goldbach: His idea was to show the existence of infinitely many **relatively prime numbers**: two numbers a, b are relatively prime if they don't have a common factor. We will show that the Fermat numbers, $F_n := 2^{2^n} + 1$, $n = 0, 1, 2, \ldots$, are relatively prime. To do that we first derive a recursive definition. Let's consider

$$F_n - 2 = (2^{2^{n-1}})^2 - 1 = (2^{2^{n-1}} + 1)(2^{2^{n-1}} - 1) = F_{n-1}(F_{n-1} - 2) = F_{n-1}F_{n-2} \ldots F_0.$$

Suppose F_n had a non-trivial common factor with F_k, $k < n$, then this common factor must divide $F_n - \prod 0 \leq i < n F_i = 2$. Thus this common factor is 2, which means that F_n is an even number, a contradiction.

¶3. Folklore: Given a prime p, the idea is to show that all the primes q dividing **Mersenne number** $2^p - 1$ are greater than p; in fact, we will show something stronger, namely p divides $q - 1$.

Let's consider the set $\mathbb{Z}'_q := \{1, \ldots, q - 1\}$; clearly 2 is an element of this set. Since q divides $2^p - 1$, $2^p \equiv 1 \mod q$. Also, p is the smallest power of 2 that has this property; since if there was a $k < p$ such that $2^k \equiv 1 \mod q$ then p must be a multiple of k, so either $k = 1$ (which can't be) or $k = p$. Our claim will thus follow if we show that $2^{q-1} \equiv 1 \mod q$.

Let $S := \{1, 2, \ldots, 2^{p-1}\}$. For each $a \in \mathbb{Z}'_q$, define the **coset** $S_a := aS$; as $a2^k \neq a2^j$, for $1 \leq k < j < p$, the size of all the cosets is p. But not all the cosets are different. In fact, we can show that for $a, b \in \mathbb{Z}'_q$, $S_a = S_b$ iff $a^{-1}b \in S$ (note a is relatively prime to q so the inverse is well define). Since $a^{-1}b = 2^k$, we have $b = a2^k \in S_a$, and $a = b2^{p-k} \in S_b$. Thus we cosets form a partitioning of $\mathbb{Z}'_q$, and each has size p, so p must divide $q - 1$.[1]

¶4. Euler: Define the function $\pi(x)$ as the number of primes not exceeding x. Then there are infinitely many primes iff $\pi(x)$ is not a decreasing function. We thus need to derive a lower bound on $\pi(x)$ that does not decrease.

Let us number the primes $p_1, p_2, p_3, \ldots$ so on. Given some number n, let $n \leq x < n + 1$. Then by comparing the area under the $\ln$ function and the Harmonic function H_n, we know that

$$\ln x \leq H_n = \sum_{k=1}^{n} \frac{1}{k} < \sum_{k} \frac{1}{k}$$

where the last sum is over all numbers k whose prime divisors are less than or equal to x. Every k can be expressed as a product of primes $p \leq x$, thus $k = \prod_{p \leq x} p^{e_p}$, where $e_p \geq 0$. But as k is varying over all numbers whose prime divisors are $\leq x$, we get

$$\ln x = \prod_{p \leq x} \sum_{e \geq 0} \frac{1}{p^e} = \prod_{p \leq x} \frac{1}{1 - 1/p} = \prod_{p \leq x} \frac{p}{p-1}. = \prod_{p \leq x} (1 + \frac{1}{p-1}) = \prod_{i=1}^{\pi(x)} (1 + \frac{1}{p_i - 1}).$$

[1] This is just a statement of Lagrange's theorem in the case of cyclic groups.

But clearly $p_i - 1 \geq i$, thus

$$\ln x \leq \prod_{i=1}^{\pi(x)} \frac{i+1}{i} = \pi(x) + 1.$$

As we very well know $\ln x$ is an increasing function, and so is $\pi(x)$.

¶5. Euler/Erdös: The last proof not only shows that there are an infinitude of primes, but it shows that their inverse sum is diverging, i.e., $\sum[p \text{ prime}]1/p$ is unbounded. If there were only finitely many primes, this wouldn't be the case. This result goes back to Euler, but the proof we present here, a proof by contradiction, is Erdös's.

Let's order the primes in increasing order $p_1, p_2, p_3, \ldots$ so on. Suppose the sum $\sum_{i\geq 1} 1/p_i$ converges, then there must be an index k such that $\sum_{i>k} 1/p_i < 1/2$. Let us call the primes $p_1, \ldots, p_k$ the **small primes** and the primes $p_{k+1}, p_{k+2}, \ldots$ the **large primes**. For an *arbitray number* N we thus have

$$\sum_{i>k} \frac{N}{p_i} < \frac{1}{2}. \tag{1}$$

Let N_b be the count of numbers $n \leq N$ that are divisble by at least one big prime, and N_s be the count of number $n \leq N$ that have only small prime divisors. Thus $N_b + N_s = N$. We will derive a contradiction to this statement. Given a prime p_i, the quantity $\lfloor N/p_i \rfloor$ counts the number of n's smaller than N that are divisble by p_i. Thus

$$N_b \leq \sum_{i>k} \left\lfloor \frac{N}{p_i} \right\rfloor \leq \sum_{i>k} \frac{N}{p_i} < N/2.$$

So to get to our contradiction, we only need to show that $N_s < N/2$. Let $n \leq N$ be a number that has only prime factors. We can factorize n as $n = ab^2$, where a is the square-free part of n, i.e., a is the product of *different* small primes. We will construct an upper bound on how many such n can be there? Since a is the product of *different* small primes, and there are only k small primes, there are at most 2^k different values of a. How many different values can b take? We don't want to apply the same argument that we had for a to b, because we have to account for the various choices of exponents that the primes can have in b. A straightforward bound on b is $b \leq \sqrt{n} \leq \sqrt{N}$. Thus $N_s < 2^k\sqrt{N}$. To get the contradiction we have to find an N, such that $2^k\sqrt{N} \leq N/2$, or $N \geq 2^{2k+2}$. Since our choice of N was arbitrary, we can choose such an N.

We have seen different proofs for the infinity of primes. How about primes modulo 4? Well, there can only be three types of primes modulo 4: those congruent to two (which is only 2), those of the form $4k+1$ and those of the form $4k+3$. But can it be that we've infinitely many primes of the last two types? Let's start with the primes of the form $4k+3$ and try to construct an argument similar to Euclid's. Suppose there were only finitely such primes, and let p_k be the largest amongst them. Then the number

$$N := 4 \cdot 3 \cdot 5 \cdots p_k - 1$$

is congruent to 3 modulo 4, and hence it has a prime factor p, distinct from $2, 3, 5, \ldots, p_k$ that is of the form $4k+3$; if all prime factors p of N are congruent to 1 modulo 4 then so is their product of powers, which would mean $N \equiv 1 \mod 4$, a contradiction.

Why are there infinitely many primes of the form $4k+1$? Let's try the same argument as above: Define $N := 4 \cdot 3 \cdot 4 \cdot p_k + 1$, where p_k is the largest prime of the form $4k+1$. Now $N \equiv 1 \mod 4$, so there must be a prime factor p distinct from $2, 3, \ldots, p_k$. But why is $p \equiv 1 \mod 4$? The eariler argument fails, because a prime $p \equiv 3 \mod 4$ on squaring gives us a number equivalent to 1 modulo 4. The proof for the infinity of primes of the form $4k+1$ is a bit complicated.

The idea is to first show that there is no solution to the equation $x^2 + 1 \equiv 0 \mod p$, where $p = 4k+3$; then we show that primes of the form $4k+3$ cannot divide numbers written as sum of two squares; and finally, we construct such a number one of whose prime factors will be of our desired form. However, the proof of this claim follows from an answer to a question that Fermat asked: Which numbers can be written as sums of two squares? We present this argument, and as a corollary we will get an answer to the question that we wanted. We then go on to show Lagrange's four square theorem, i.e., every number can be expressed as sum of at most four squares.

Lagrange's theorem goes beyond Fermat's and states that *every number can be represented as sum of squares of four integers*. If we want to replace integers with natural numbers then the statement would be square of *at most* four natural numbers, which is not as neat as the previous formulation. Here we derive a proof using geometry of numbers.