The Distribution of Prime Numbers

Ryan C. Daileda

Trinity University

Number Theory

Introduction

The definition of a prime number is simple and intuitive.

And, with the possible exception of Euclid's Lemma, the proof of the FTA is also relatively straightforward.

Despite this, the prime numbers themselves are elusive. There are many easily observed patterns that defy explanation.

Today we will take a tour of results concerning the primes: their distribution, their spacing, and special forms that occur.

Recall the prime counting function:

$$\pi(x) = \#\{p \leq x \mid p \text{ is prime}\}.$$

The infinitude of primes (due to Euclid) can be rephrased as

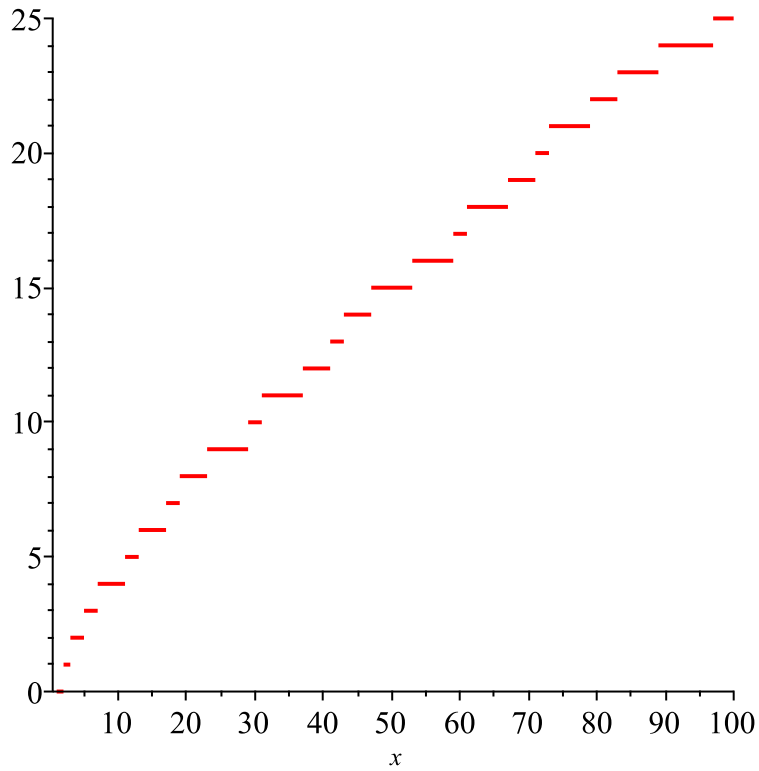
$$\lim_{x \to \infty} \pi(x) = \infty.$$

A more precise result is the *Prime Number Theorem* (PNT):

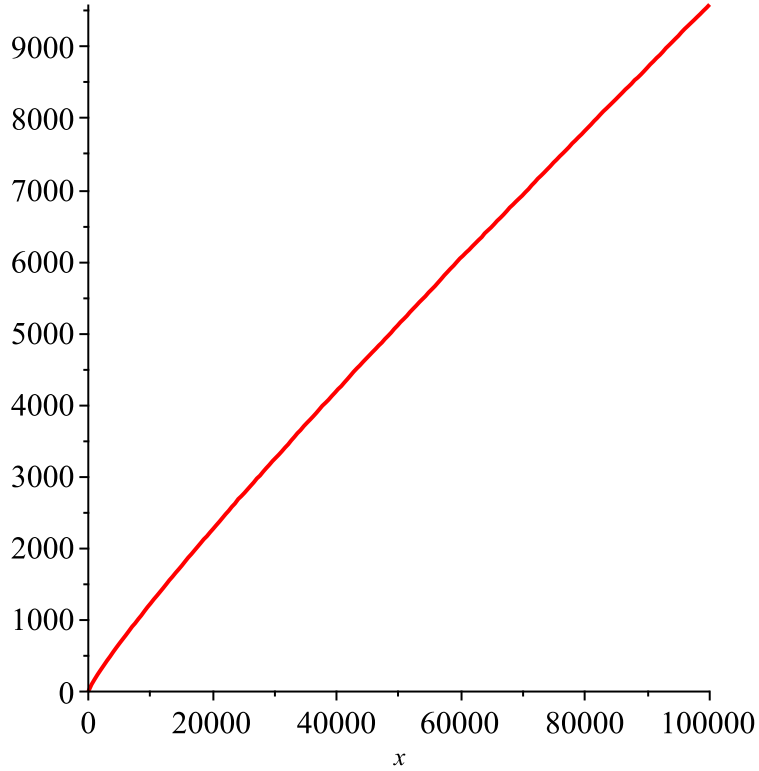
$$\lim_{x \to \infty} \frac{\pi(x)}{x/\log x} = 1.$$

Conjectured by Gauss and others around 1800, it wasn't until 1896 that Hadamard and de la Vallée Poussin (independently) succeeded in proving the PNT.

Although it has discontinuities at every prime, on large scales the graph of $\pi(x)$ appears remarkably smooth.



$\pi(x)$ for $1 \leq x \leq 100$



$\pi(x)$ for $1 \leq x \leq 10^5$

The Riemann Hypothesis

The *Riemann Zeta Function* is defined by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

for $\mathrm{Re}(s) > 1$.

$\zeta(s)$ can be immediately connected to the prime numbers via its *Euler product expansion*,

$$\zeta(s) = \prod_{p} \left(1 - \frac{1}{p^s}\right)^{-1},$$

which follows from the FTA.

$\zeta(s)$ can be analytically continued to $\mathbb{C} \setminus \{1\}$, and it satisfies a *functional equation* relating its values at s and $1 - s$.

The behavior of the zeros of $\zeta(s)$ in the complex plane influence the behavior of the primes.

Indeed, one has the explicit formula

$$\sum_{p^m \leq x} \log p = x - \sum_{\rho} \frac{x^\rho}{\rho} - \log 2\pi - \frac{1}{2} \log(1 - x^{-2}),$$

where the sum on the left runs over all the prime powers p^m up to x, and the sum on the right runs over the zeros ρ of $\zeta(s)$ in the *critical strip* $0 < \mathrm{Re(s)} < 1$.

The *Riemann hypothesis* asserts that $\mathrm{Re}(\rho) = \frac{1}{2}$ for all ρ, and gives very sharp estimates on the error term in the PNT.

The are a number of more "refined" questions one can pose about the distribution of the primes.

Theorem 1 (Bertrand's Postulate)

For any $n \geq 2$, there is a prime p satisfying $n \leq p < 2n$.

Remarks.

- Conjectured by Bertrand in 1845, this result was first proven by Chebyshev in 1852.
- Bertrand's Postulate gives a (rather weak) bound on the growth of the nth prime number: $p_n < 2^n$.
- Ramanujan and Erdős gave simpler proofs in the twentieth century.

Distances Between Primes

The main questions here are:

- How close together can two (successive) primes be?
- How often?

Because every prime $p > 2$ is odd, we must have

$$p_{n+1} - p_n \geq 2$$

for $n \geq 2$.

The *Twin Primes Conjecture* asserts that $p_{n+1} - p_n = 2$ infinitely often.

If we let

$$\pi_2(x) = \{p \leq x \,|\, p \text{ and } p+2 \text{ are prime}\},$$

the Hardy-Littlewood conjecture is the asymptotic

$$\pi_2(x) \sim C \int_2^x \frac{dt}{(\log t)^2},$$

for a certain constant C.

The best known result to date (due to Zhang, Tao and others) is:

Theorem 2

There exist infinitely many n for which

$$p_{n+1} - p_n \leq 246.$$

A related question is how far apart successive primes can be. Here there is a definitive (and elementary) answer.

Theorem 3

Let $n \in \mathbb{N}$. Then there is a k so that

$$p_{k+1} - p_k \geq n.$$

Proof. This is equivalent to the assertion that, given $n \in \mathbb{N}$, there exist n consecutive composite numbers.

Let $n \in \mathbb{N}$ and consider the n consecutive integers

$$(n+1)! + 2, (n+1)! + 3, (n+1)! + 4, \ldots, (n+1)! + (n+1).$$

Since k divides both $(n+1)!$ and k, it divides $(n+1)! + k$.

Since $1 < k < (n+1)! + k$, this proves each $(n+1)! + k$ is composite. □

Remark. Notice that we have two contrasting situations concerning gaps between successive primes:

- $p_{n+1} - p_n$ is (conjecturally) as small as possible infinitely often.
- $p_{n+1} - p_n$ can be made arbitrarily large.

Primes of Special Form

Question. Given a function $f : \mathbb{N} \to \mathbb{N}$, how often is $f(n)$ prime?

Examples.

- Primes of the form $2^n - 1$ (i.e. we take $f(n) = 2^n - 1$) are called *Mersenne primes*.
- Integers of the form $2^{2^n} + 1$ are called *Fermat numbers*.

The largest explicitly known prime number is a Mersenne prime:

$$p = 2^{77232917} - 1.$$

It is not hard to prove:

Theorem 4

If $2^n - 1$ is prime, then n is prime.

So all Mersenne primes have prime exponents.

While this narrows down the candidates for Mersenne primes, it doesn't guarantee their existence.

Conjecture. There are infinitely many Mersenne primes.

Concerning the Fermat numbers:

- Fermat conjectured *every* integer of the form $2^{2^n} + 1$ is prime.
- $2^{2^n} + 1$ is prime for $n = 0, 1, 2, 3, 4$.
- $2^{2^n} + 1$ is known to be *composite* for $5 \leq n \leq 32$.

Primes Given By Polynomials

Question: Given a polynomial $f(X) \in \mathbb{Z}[X]$, how often is $f(n)$ prime?

Restrictions need to be placed on $f(X)$ in order to avoid trivial results:

- $f(X)$ should be irreducible.
- The coefficients of $f(X)$ should be relatively prime.

Even with these restrictions, however, one can show that:

Theorem 5

If $f(X) \in \mathbb{Z}[X]$ is nonconstant, then there exist infinitely many $n \in \mathbb{N}$ for which $f(n)$ is composite.

An easy proof can be given using modular arithmetic.

If $\deg f > 1$, not much is known, even in very simple cases. A famous example is:

Conjecture. There are infinitely many primes of the form $n^2 + 1$.

On the other hand, if $f(n) = an + b$ (i.e. f is linear), a great deal can be said.

Definition

An *arithmetic progression* is a set of the form

$$\{an + b \,|\, n \in \mathbb{Z}\} = \{\ldots, b - a, b, b + a, b + 2a, b + 3a, \ldots\},$$

where $a, b \in \mathbb{Z}$ and $a \neq 0$.

Using tools from analysis, in 1837 Dirichlet proved:

Theorem 6 (Dirichlet's Theorem on Primes in Progressions)

Every arithmetic progression with $(a, b) = 1$ *contains infinitely many primes.*

Equivalently, if $(a, b) = 1$, then the polynomial $f(n) = an + b$ is prime infinitely often.

The PNT also generalizes to primes in progressions. Specifically, let

$$\pi_{a,b}(x) = \#\{p \leq x \mid p \text{ is prime and } p = an + b\}.$$

Then it can be shown that

$$\lim_{x \to \infty} \frac{\pi_{a,b}(x)}{x/\log x} = \frac{1}{\varphi(a)},$$

where φ denotes Euler's totient function.

There is no known purely arithmetic proof of Dirichlet's theorem, in general.

In certain special cases, however, it can be established by generalizing Euclid's proof of the infinitude of primes.

Note that every odd prime necessarily has one of the forms $4n + 1$ or $4n + 3$, by the division algorithm.

Let's consider the case of primes of the form $4n + 3$.

Lemma 1

If $a, b \in \mathbb{N}$ both have the form $4n+1$, then so does ab.

Proof. Write $a = 4m+1$, $b = 4n+1$ with $m, n \in \mathbb{Z}$. Then

$$ab = (4m+1)(4n+1) = 16mn+4m+4n+1 = 4(4mn+m+n)+1.$$

Since $4mn + m + n \in \mathbb{Z}$, this proves the result. □

Theorem 7

There are infinitely many primes of the form $4n+3$.

Proof. Suppose, for the sake of contradiction, that there are only finitely many such primes:

$$p_1, p_2, \ldots, p_n.$$

Let

$$N = 4p_1 p_2 \cdots p_n - 1.$$

Notice that N has the form $4n + 3$:

$$N = 4p_1 p_2 \cdots p_n - 1 = 4(p_1 p_2 \cdots p_n - 1) + 3.$$

Since N is odd, every one of its prime factors has the form $4n + 1$ or $4n + 3$.

If every prime factor of N had the form $4n + 1$, so would N, by the lemma.

So there is a prime of the form $4n + 3$ that divides N.

That is, $p_i | N$ for some i. But then

$$p_i | N - 4p_1 p_2 \cdots p_n = -1,$$

which is impossible. □

Primes Represented by Quadratic Forms

Definition

A *quadratic form (in two variables)* is a homogeneous quadratic polynomial $Q(x, y) = ax^2 + bxy + cy^2$ with $a, b, c \in \mathbb{Z}$.

For example, the simplest quadratic form is $Q(x, y) = x^2 + y^2$.

Definition

Let $Q(x, y)$ be a quadratic form and $n \in \mathbb{Z}$. We say that $Q(x, y)$ *represents* n if there exist $a, b \in \mathbb{Z}$ so that $Q(a, b) = n$.

Question. Given a quadratic form $Q(x, y)$, which primes does it represent?

Theorem 8 (Fermat)

A prime p is represented by the quadratic form $x^2 + y^2$ if and only if p has the form $4n + 1$.

For example, $29 = 4 \cdot 7 + 1$ and we have

$$29 = 2^2 + 5^2.$$

Along the same lines we have the following result, originally conjectured by Euler:

Theorem 9

A prime p is represented by the quadratic form $x^2 + 5y^2$ if and only if p has the form $20n + 1$ or $20n + 9$.

For instance, $29 = 20 \cdot 1 + 9$ and we have

$$29 = 3^2 + 5 \cdot 2^2.$$

The question of the representability of primes by quadratic forms of the type $x^2 + ny^2$ has been completely settled, and requires deep ideas from *algebraic number theory.*