PRIMALITY TESTING

Efrain Alvarado Jeremy Bates Duc Cao Sergio Lara Xuan Duc Tran
University of Illinois at Chicago
MSCS Undergraduate Research Laboratory

About this project

This project was supervised by Evangelos Kobotis. There were weekly zoom meetings in which we went over theoretical points, set goals and discussed our results.

Motivation

The general purpose of this project was to explore the theory of prime numbers through a series of numerical experiments. Our main goal was to identify large prime numbers and to go over the different means of testing primality.

Elementary approach and choice of range

Our first task was to produce a list of prime numbers. To this end we came up with all prime numbers that do not exceed 10^8. This can be achieved by using quite classical means in a very small amount of time. This allowed us to start a further programming task by generating the primes in this range and then manipulating them as desired. In this way we were able to look for special kinds of primes, analyze their distribution and in general get acquainted with the speed that it takes to solve numerical problems like this in the range that we ended up referring to as the *trivial range*.

Our intention was to move outside the trivial range and explore things beyond its confines. By having all prime numbers within our trivial range we were able to find prime numbers in what we called the *accessible range*. This is the range of numbers that do not exceed 10^{16}. In particular we were able to easily examine ranges consisting of 10^6 numbers and determine all of their primes.

Additional exploration

Simultaneously we sought to connect our results with theoretical facts such as those provided by the Prime Number Theorem. In particular we were able to compare the amount of prime numbers in a given range by those projected by the Prime Number Theorem. We were always conscious about the approximate nature of such a comparison but it turned out that our computations were consistent with everything that the theory predicted.

Methodology

During our meetings everyone was engaged, bringing new ideas, discussing their successes and failures with the assignments of the previous week, volunteering for the new tasks or giving suggestions for different directions. The role of this presentation is to summarize some of the work that we did. In many cases we took code segments from the web but overall the code and results that we generated were our own work. We made every effort to give credit to our outside sources.

Starting with an ancient algorithm

Exploring the properties of the set of prime numbers is one of the most fascinating and difficult tasks in mathematics. We began by looking over an ancient algorithm that theoretically produces all prime numbers. This is the Eratosthenes sieve and it is based upon the fact that every composite natural number is divisible by a prime number that does not exceed its square. Indeed if n is composite, then it can be written as ab where a and b are natural numbers greater than 1. If $a \geq b$, then $b^2 \leq n$ and if p is any prime number dividing b then $p^2 \leq n$ or $p \leq \sqrt{n}$.
This simple property has the following implication. If we start with the set:

$$\{1,2,3,\ldots,k^2\}$$

and we strike out 1 and all the numbers that are divisible by primes not exceeding k, then the numbers that are left behind are precisely the primes that are between k and k^2. In other words, if we know the primes that are less than k then we can easily find the primes that are less than k^2. More concretely, if we begin with the primes $2,3,5,7$ that are precisely the primes that do not exceed 10, then we can easily find the primes that do not exceed 100. Once we can accomplish that, then we can easily find the primes not exceeding 10,000 and so on. We did find 5,761,455 primes less than 10^8.

The Prime Number Theorem

The Prime Number Theorem was conjectured by Carl Friedrich Gauss and Jean-Marie Legendre. In modern terms, it states that if we set:

$$\pi(x) = \#\{p \text{ prime} : p \leq x\}$$

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

Primality testing

Finding large prime numbers is one of the main goals in the theory of prime numbers. Here we apply a simple-minded search based on several theoretical results that we mention below. There are several primality tests. Some of them are probabilistic. Other are deterministic. Probabilistic primality tests identify numbers that have a strong probability of being prime. Deterministic tests, prove, when certain conditions are satisfied, that a given number is prime.

For some of this tests, it is useful to know Fermat's little theorem according to which if p is prime then for any integer a not divisible by p, we have:

$$a^{p-1} \equiv 1 \mod p$$

This means that if we have a number n for which the congruence:

$$a^{n-1} \equiv 1 \mod n$$

Primality testing (continued)

is wrong for a given a that is coprime to n, then it cannot be a prime number. On the other hand, if this is correct for a given a which is coprime to n, then this may make us hope (but certainly not decide) that n is prime. In fact this is the content of the **Fermat Primality Test**. One chooses a number a less than n and tests the equality $a^{n-1} \equiv 1 \mod n$. If it is true then we think of n as having some probability of being prime.

The **Miller-Rabin test** is also probabilistic and it is based on the following process. We have a number n and we consider a number a which is coprime to n. We then write $n-1 = 2^s m$, where m is an odd number. We then test to see if the numbers

$$a^m, a^{2m}, \ldots, a^{2^s m}$$

are all equal to 1 with the possible exception of the first one that could be -1. If this is the case, then the test is passed by n and it has a strong probability to be a prime number. At this point, it should be noted that the Miller-Rabin test is, in fact, deterministic as long as the Generalized Riemann Hypothesis is correct. From this perspective, and if one believes this hypothesis, then the Miller-Rabin test can be used as a deterministic test.

We however explored another test which is known to be deterministic. This is the Agrawal-Kayal-Saxena primality test, known as the **AKS primality test**. It is based upon the fact that for a natural number $n > 1$ and a coprime to n, then the polynomial relation:

$$(X+a)^n \equiv X^n + a \mod n$$

is true. This can be thought of as a relation in the polynomial ring $\mathbb{Z}/n\mathbb{Z}[X]$.

Finally, it is conjectured that the following test is deterministic. If n is natural number satisfying:

- $2^{n-1} \equiv 1 \mod n$
- $F_{n+1} \equiv 0 \mod n$

then n is a prime number. Here F_{n+1} is the $n+1$-th term of the Fibonacci sequence defined by $F_0 = F_1 = 1$ and $F_{n+2} = F_{n+1} + F_n$.

Numerical results in the accessible range

Let's begin by taking a look at the interval $[10^{10}, 10^{10} + 10^6]$. The prime number theorem projects that there are going to be 43,429 primes in this range. We begin by sieving out all the multiples of primes in the trivial range (recall that the trivial range consists of all numbers up to 10^8). Then we find 43,427 primes: a number which is suprisingly close to what is predicted by the Prime Number Theorem.

If we repeat this in the interval $[10^{11}, 10^{11} + 10^6]$ then we find 39,434 primes compared to 39,484 primes that are predicted by the Prime Number Theorem. Note that the proportion of primes is decreasing as we go further into larger and larger numbers.

Moving further

In the accessible range (below 10^{16}), things can be done in a rather straightforward way by very elementary means. The question is what happens when we go much further. Let's take a look at the interval $[10^{50}, 10^{50} + 10^5]$.

We have 100,000 numbers to test in this range. We start by striking out all the numbers that are multiples of primes that do not exceed 10^8. This leaves us with 4,064 numbers.

We next use the Fermat primality test and strike out almost 90% of the above numbers to come up with 895 numbers that have a good probability of being prime numbers. In other words we have less than 1% of the numbers we started up with. This is a manageable amount of numbers to which we can apply further tests including the conditionally deterministic Miller-Rabin test. This allows us to bring the number down to 452. These 452 numbers pass all the different primality tests and give us the final amount of primes in the range that we considered. Note that it takes a little over a minute to run these computations. One should note here that throughout our project we tried to keep the computations as short as possible.

By going even further, we do the same thing in the range $[10^{400}, 10^{400} + 10^5]$, dealing with some seriously large numbers. It comes as an initial surprise that the initial sieving produces 4,055 numbers that have some probability of being primes. This is an amount that is almost equal to what we found at a much smaller range. The reason this might be at first surprising is because we know that prime numbers do become more and more scarce. However we do not improve on the set of primes by which we are sieving and therefore we should not expect any particular improvement on the amount of numbers that we strike out this way. It is in fact true that we found about 4,000 numbers in every range of length 100,000 where we applied this sieve.

What however becomes much more interesting (and consistent with theoretical expectations) is the fact that the subsequent tests strike out significantly more numbers. In fact we are left with 66 numbers that are prime. 66 numbers out of 100,000 consecutive numbers are prime. This process takes about 8 minutes and the larger prime that is produced this way is the number:

100000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000000000
000000000000000000000000000000000000000000000041977

which is actually the largest prime number that we found with our computations. Allowing more computation time, would have enabled us to go even further. However it is already impressive that our simple-minded techniques took us to computations that would have been impossible only a short period of time ago, let alone that it takes only a few minutes to complete.

Looking back at our journey in this project, one can easily come up with future goals that involve more precise computations in higher number ranges along with comparisons to theoretical data.