Abstract

In this paper I will cover how to calculate how frequently a prime is the smallest divisor over an interval. I will then extend that definition to show how frequently that prime is the smallest divisor of all integers. Lastly , I will use this information to discuss very composite numbers and very prime numbers .

Key Definitions

1. **Coprime (Relatively Prime) Integers:** Two integers a and b are coprime if their greatest common divisor is 1:

$$\gcd(a, b) = 1.$$

2. **Primorial :** The primorial of p_n, denoted P_n, is the product of the first n primes :

$$P_n = \prod_{p \leq p_n} p.$$

3. **Totient Function :** The Euler totient function $\phi(a)$ is calculated using the prime factorization of a. If a has the prime factorization:

$$a = p_1^{e_1} p_2^{e_2} \cdots p_b^{e_b},$$

then the totient function is given by:

$$\phi(a) = a \prod_{i=1}^{b} \left(1 - \frac{1}{p_i}\right),$$

where $p_1, p_2, \ldots, p_b$ are the distinct prime factors of a. This formula represents the count of integers $x < a$ that do not share any factors with a .

Using the Totient Function to Determine the Frequency of a Prime as the Smallest Factor Over a Fixed Interval

$P_n = a \cdot p_n$

Consider the product $a \cdot p_n$, where $a = p_1 \cdot p_2 \cdot \cdots \cdot p_{n-1}$ is the product of the first $n-1$ primes , and p_n is the n-th prime . We wish to determine the number of times p_n is the smallest factor of integers in the interval $[1, a \cdot p_n]$.

Integers Factorable by p_n

All integers in $[1, a \cdot p_n]$ that are divisible by p_n take the form:

$$p_n \cdot 1,\ p_n \cdot 2,\ p_n \cdot 3,\ \ldots,\ p_n \cdot a.$$

This set contains a elements, as a represents the range of possible values for the multiplier m.

Integers divisible by no smaller primes than p_n

For p_n to be the smallest factor of an integer $p_n \cdot m$, m must not share any factors with a, if m shares a factor with a , then m is divisible by a prime less than p_n , since a is composed of all the primes less than p_n. The number of integers j in $[1, a]$ that are coprime to a is given by $\phi(a)$. Thus, there are $j = \phi(a)$ integers in $[1, a \cdot p_n]$ which take the form $p_n \cdot m$, where p_n is the smallest factor . The totient function $\phi(a)$ allows us to determine the exact number of times p_n is the smallest factor of integers in the interval $[1, a \cdot p_n]$. This is precisely $\phi(a)$, as it counts the values of m coprime to a .

Frequency of p_n as the Smallest Factor

To compute the relative frequency in $[1, a \cdot p_n]$ where p_n is the smallest factor:

$$\text{Frequency} = F_n = \frac{\phi(a)}{a \cdot p_n} = \frac{\phi(P_{n-1})}{P_n}$$

Example 1

Let $n = 2$ then $p_2 = 3$. Let a be the product of all primes less than $p_2 = 3$. $a = 2$

$$\phi(a) = \phi(2) = 2 \cdot (1 - \frac{1}{2}) = 2 \cdot \frac{1}{2} = 1$$

$$a \cdot p_n = 2 \cdot 3 = 6$$

$$F_2 = \frac{\phi(a)}{a \cdot p_n} = \frac{1}{6}$$

We can say that $p_2 = 3$ is the smallest divisor of $\frac{1}{6}$ of all integers . By use of the totient function we may exclude m where m is a factor of a .

$$3 \cdot 1, \cancel{3 \cdot 2},$$

Example 2

Let $n = 3$ then $p_3 = 5$. Let a be the product of all primes less than $p_3 = 5$. $a = 2 \cdot 3 = 6$

$$\phi(a) = \phi(6) = 6 \cdot (1 - \frac{1}{2}) \cdot (1 - \frac{1}{3}) = 6 \cdot \frac{1}{2} \cdot \frac{2}{3} = 2$$

$$a \cdot p_n = 6 \cdot 5 = 30$$

$$F_3 = \frac{\phi(a)}{a \cdot p_n} = \frac{2}{30}$$

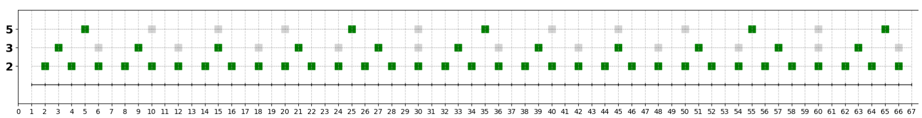
We can say that $p_3 = 5$ is the smallest divisor of $\frac{2}{30}$ of all integers .By use of the totient function we may exclude m where m is a factor of a .

$$5 \cdot 1, \cancel{5 \cdot 2}, \cancel{5 \cdot 3}, \cancel{3 \cdot 4}, 5 \cdot 5, \cancel{5 \cdot 6},$$

Case for F_1 , $p_1 = 2$

F_1 is a special case since there are no primes smaller than 2 . So the frequency for 2 will be calulated heuristically . Since 2 divides $\frac{1}{2}$ of all integers and there are no primes smaller than 2 we will say that F_1 is $\frac{1}{2}$. 2 is the smallest divisor of $\frac{1}{2}$ of all integers .

Visual Proof



Green ticks are when p_n is the smallest factor of an integer . Grey ticks are for when p_n divides an integer but there are smaller primes which also divide that integer . We can see that the frequency of p_n to divide an integer as the smallest factor, repeats over the interval of its primorial . This is because frequency is measured relative to smaller primes and its primorial is the least common factor $a \cdot p_n$.

Sum of frequencies

The total sum of frequencies S_n of integers in $[1, P_n]$ that are divisible by some prime p_i up to p_n, for which p_i is the smallest factor, is given by:

$$S_n = \sum_{i=1}^{n} F_i = \sum_{i=1}^{n} \frac{\phi(P_{i-1})}{P_{i-1} \cdot p_i} = \sum_{i=1}^{n} \frac{\phi(P_{n-1})}{P_n}.$$

S_n is the measure of the likelihood of an integer to share a factor with a prime less than or equal to p_n . The likelihood that an integer is coprime to S_n is $1 - S_n$

Approximating the behavior of F_n for large n

We start from

$$F_n = \frac{\phi(P_{n-1})}{P_n}.$$

$$F_n = \frac{\phi(P_{n-1})}{P_n} = \frac{P_{n-1} \prod_{i=1}^{n-1} \left(1 - \frac{1}{p_i}\right)}{P_{n-1} \cdot p_n}.$$

$$F_n = \frac{\cancel{P_{n-1}} \prod_{i=1}^{n-1}\left(1-\frac{1}{p_i}\right)}{\cancel{P_{n-1}} \cdot p_n} = \frac{\prod_{i=1}^{n-1}\left(1-\frac{1}{p_i}\right)}{p_n}.$$

We can use Mertens' third theorem aka Mertens' product formula , to approximate the numerator asymptotically , we have

$$\prod_{i=1}^{n-1}\left(1-\frac{1}{p_i}\right) \sim \frac{e^{-\gamma}}{\log p_{n-1}}.$$

where γ is the Euler–Mascheroni constant. This provides us with an asymptotic approximation for the numerator .
Since $p_{n-1} \approx p_n$ for large n, this gives

$$\prod_{i=1}^{n-1}\left(1-\frac{1}{p_i}\right) \approx \frac{e^{-\gamma}}{\log p_n}.$$

To complete the original definition we must factor in the denominator .

$$F_n = \frac{1}{p_n}\prod_{i=1}^{n-1}\left(1-\frac{1}{p_i}\right) \approx \frac{1}{p_n} \cdot \frac{e^{-\gamma}}{\log p_n}.$$

Given

$$F_n \approx \frac{e^{-\gamma}}{p_n \log p_n},$$

and $p_n \sim n \log n$, we have

$$F_n \approx \frac{e^{-\gamma}}{n \log n \cdot \log(n \log n)}$$

$$F_n \approx \frac{e^{-\gamma}}{n \log n \cdot (\log n + \cancel{\log \log n}))}$$

For large n F_n will grow as $\log n$ as opposed to $\log \log n$, since $\log n \gg \log \log n$ for large n .

$$F_n \approx \frac{e^{-\gamma}}{n(\log n)^2}$$

Thus:

$$S_n \sim \sum_{h=1}^{n} \frac{e^{-\gamma}}{h(\log h)^2}$$

As $n \to \infty$, $F_n \to 0$.

Extremely composite integers and extremely coprime integers

Let P_n be the product of the first n primes $G = \{2, 3, 5, ..., p_n\}$. This means P_n is factorable by all of these primes . Thus P_n is composite with respect to S_n. It can be shown that $P_n \pm 2$ is factorable by 2 , $P_n \pm 3$ is factorable by 3 , $P_n \pm 5$ is factorable by 5 etc . It can be deduced that $P_n \pm 1$ is not factorable by any of the elements in G . $P_n \pm 1$ is coprime to G .

Error Terms

There are primes $p_{n+1}, p_{n+2}, ...$ in the interval $[p_n, P_n]$ whose frequencies are not accounted for and which may be a factor of $P_n \pm 1$, we will call the sum of the frequencies of these primes E_n. Let K be the set of primes accounted for in E_n. If $P_n \pm 1$ is coprime to the primes used to calculate S_n, G and coprime to the primes used to approximate E_n, K then $P_n \pm 1$ is prime .

Quantifying the Error Term E_n

The upper limit $\sqrt{P_n + 1}$ is chosen because $\sqrt{P_n + 1}$ is the largest possible factor which can divide $P_n + 1$. We have:

$$E_n = \sum_{m>n}^{\sqrt{P_n+1}} F_m,$$

To estimate E_n, consider the sum:

$$E_n = \sum_{m>n} \frac{e^{-\gamma}}{m(\log m)^2}.$$

For large n, we can approximate sums by integrals. Specifically:

$$\sum_{m>n} \frac{1}{m(\log m)^2} \approx \int_n^\infty \frac{dx}{x(\log x)^2}.$$

Simplify and Evaluate the Integral

$$u = \log x$$

$$\frac{du}{dx} = \frac{d}{dx}(\log x) = \frac{1}{x} \quad \Rightarrow \quad du = \frac{1}{x}dx \quad \Rightarrow \quad dx = x\,du$$

$$\frac{1}{x(\log x)^2} = \frac{1}{xu^2}$$

$$I = \int_n^\infty \frac{dx}{xu^2} = \int_{\log n}^\infty \frac{x\,du}{xu^2} = \int_{\log n}^\infty \frac{du}{u^2}$$

$$\int \frac{du}{u^2} = \frac{1}{u^2} = u^{-2} = \frac{u^{-2+1}}{-2+1} = \frac{u^{-1}}{-1} = -\frac{1}{u} + C$$

$$I = \left[-\frac{1}{u}\right]_{\log n}^\infty = \left(-\frac{1}{\infty}\right) - \left(-\frac{1}{\log n}\right) = 0 - \left(-\frac{1}{\log n}\right) = \frac{1}{\log n}$$

Thus:

$$\int_n^\infty \frac{dx}{x(\log x)^2} = \frac{1}{\log n}.$$

Incorporating the constant $e^{-\gamma}$:

$$E_n \approx e^{-\gamma} \cdot \frac{1}{\log n} \quad \text{as } n \to \infty.$$

E_n is on the order of $\frac{1}{\log n}$. If the sum defining E_n is not from n to ∞ but to some finite upper limit M we can write:

$$\int_n^M \frac{dx}{x(\log x)^2} = \left(-\frac{1}{\log M}\right) - \left(-\frac{1}{\log n}\right).$$

If M is extremely large the dominant term is $1/\log n$, ensuring it remains on the order of $1/\log n$ for large n.

$$E_n \approx \frac{e^{-\gamma}}{\log n}.$$

$$E_n \to 0 \text{ as } n \to \infty$$

Extremely prime integers

Let P_n be the product of the first n primes $G = \{2, 3, 5, ..., p_n\}$. $P_n \pm 1$ is not factorable by any of the elements in G . $P_n \pm 1$ is coprime to G . Let K be the primes accounted for in E_n. As n increases , the likelihood of $P_n \pm 1$ being divisible by K decreases as $n \to \infty$ since as $n \to \infty, E_n \to 0$. It follows that as $n \to \infty, P_n \pm 1$ has an increasing likelihood of being prime . The likelihood that either $P_n \pm 1$ is prime is approximately $(1 - E_n)$. So the likelihood that both are prime is approximately $(1 - E_n)^2$. Therefore, as n increases, the likelihood that $P_n \pm 1$ are prime increases.

Proof

Let A denote the infinite ordered set of all prime numbers. We construct a subset $B \subseteq A$ consisting of primorials that serve as twin prime centers. The construction proceeds iteratively as follows:

1. For each positive integer n, define the primorial P_n as the product of the first n primes:
$$P_n = \prod_{i=1}^{n} p_i,$$
where p_i is the i-th prime number.

2. Define B as the set of primorials P_n for which both $P_n - 1$ and $P_n + 1$ are prime. Formally,
$$B = \left\{ P_n \in A \;\middle|\; P_n - 1 \text{ and } P_n + 1 \text{ are both prime} \right\}.$$

3. Analyze the likelihood that P_n is a twin prime center as n increases.

4. Define an error term E_n representing the sum of frequencies of primes beyond p_n that may divide $P_n \pm 1$:
$$E_n = \sum_{m>n}^{\sqrt{P_n+1}} F_m \approx \frac{e^{-\gamma}}{\log n}.$$
This approximation indicates that E_n diminishes as n increases.

5. Consider the likelihood that $P_n \pm 1$ is coprime to the set of primes accounted for in E_n. This is approximately $(1 - E_n)^2$. As $E_n \to 0$ when $n \to \infty$, this likelihood approaches 1. Thus $P_n \pm 1$ are increasingly likely to be prime .

6. Therefore, as n increases, the likelihood that P_n serves as a twin prime center increases. Consequently, the set B is expected to contain more elements as n grows.

7. Since B is constructed by iteratively including primorials that satisfy the twin prime center condition, and the likelihood of inclusion increases with n, we infer that B is an infinite subset of A.

8. Finally, assuming that there are infinitely many primes and that the likelihood of primorials being twin prime centers tends to 1, it follows that there are infinitely many twin prime centers.