Calculating the relative density that a prime is the smallest prime factor

Daniel Eid

December 2024

Abstract

In this paper I will cover how to calculate how frequently a prime is the smallest prime factor in some interval. I will use that information to approximate how likely specific integers near P_n are to be prime. I will show the likelihood of integers near P_n to be prime increases for large n.

Objective

The motivation for these results is to have a counting method that includes all positive integers only once. The chosen approach requires counting relative to the primes which leads to unique factorization. Given the fundamental theorem of arithmetic we know that all positive integers can be written as a unique product of primes. Each positive integer will have only one prime as its smallest divisor. This will be the premise for unique factorization. We will calculate how many times some prime is the smallest factor of integers in some interval. We will express this count in terms of n using an approximation from Mertens third theorem. We will give its count relative to some interval to determine a relative density. Finally, we will analyze sums of the first n terms and sums of the trailing terms, to determine growth and decay behavior of these prime divisibility densities.

$(p_n)_{n\in\mathbb{N}}$ is the sequence of prime numbers in ascending order

p_n **:** Let p_n be the nth prime. $n = 1, p_1 = 2$

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

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

$$P_4 = \prod_{p\leq 7} p = 7\cdot 5\cdot 3\cdot 2 = 210.$$

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 .

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

Let e be Eulers constant $e \approx 2.718$
Let γ be the Euler-Mascheroni constant $\gamma \approx 0.577$
$D_n = \frac{\phi(P_{n-1})}{P_n}$ yields the density that p_n is the smallest prime factor.
$p = 3, P_1 = 2, P_2 = 6$. $D_2 = \frac{\phi(2)}{6} = \frac{1}{6}$. 3 is the smallest prime factor of $\frac{1}{6}$ of all integers.

Main Results

1. Prime Divisibility Theorem Assume p_n , then $D_n = \frac{\phi(P_{n-1})}{P_n}$ yields the average density that p_n is the smallest prime factor.
2. Approximation Assume $D_n = \frac{\phi(P_{n-1})}{P_n}$, then $D_n \approx \frac{e^{-\gamma}}{n(\ln n)^2}$
4. Divisibility Decay Theorem Assume $E_n = \sum_{m>n} \frac{e^{-\gamma}}{m(\ln m)^2}$ is the tail sum of D_n, then $E_n > E_{n+1}$.
Section 6. Assume we have $P_n \pm 1, P_{n+1} \pm 1$, then the likelihood that $P_n \pm 1$ is prime, is less than the likelihood that $P_{n+1} \pm 1$ is prime.

1 Prime Divisibility Theorem

Assume p_n , then $D_n = \frac{\phi(P_{n-1})}{P_n}$ yields the average density that p_n is the smallest prime factor.

1.1 Determine the number of times a Prime is 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 prime factor of integers in the interval $[1, a \cdot p_n]$.
Let $n = 3$ then $p = 5$. a is the product of all primes less than $p = 5$.

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

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

1.2 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.

1.3 Integers divisible by no smaller primes than p_n

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

$$\gcd(a, b) = 1, \gcd(7, 3) = 1.$$

For p_n to be the smallest prime factor of an integer $p_n \cdot m$, m must not share any prime factors with a, if m shares a prime 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 $t = \phi(a)$ integers in $[1, a \cdot p_n]$ which take the form $p_n \cdot m$, where p_n is the smallest prime factor . The totient function $\phi(a)$ allows us to determine the exact number of times p_n is the smallest prime 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 .
Use the totient function to exclude any m divisible by a prime factor of a , when $a = 6$ and $p = 5$.

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

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

5 is the smallest prime factor 2 times in the interval $[1, 30]$.

1.4 Density of p_n as the Smallest Factor

To compute the relative density in $[1, a \cdot p_n]$ where p_n is the smallest prime factor, we will use the ratio of the number of times that p_n is the smallest prime factor, to the interval of its primorial P_n:

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

For $p = 5$ and $a = 6$

$$D_3 = \frac{\phi(a)}{a \cdot p_n} = \frac{\phi(6)}{6 \cdot 5} = \frac{\phi(6)}{30} = \frac{2}{30}$$

We can say that $p = 5$ is the smallest prime factor of $\frac{2}{30}$ of integers in the interval $[1, 30]$.

1.5 Case for D_1 , $p_1 = 2$

D_1 is a special case since there are no primes smaller than 2 . So the density for 2 will be calulated heuristically . 2 divides $\frac{1}{2}$ of the integers in the interval $[1, 2]$. There are no smaller primes than 2, so there are no smaller primes to exclude, so any time 2 divides some integer it will be the smallest prime dividing that integer. $D_1 = \frac{1}{2}$, 2 is the smallest prime factor of $\frac{1}{2}$ of integers.

1.6 Limit Behavior of D_n for large intervals

We have defined D_n for the interval up to P_n. So far it has only been defined for m of the form $P_n \cdot r + m \cdot p_n$ and $r = 0$. We must now define D_n for all $r \geq 0$ and up to some positive integer j. We show that the limit exists as $j \to \infty$. We will start by proving that the limit exists, and in a later section, show what the value of the limit approaches as $j \to \infty$.
Let $S \subseteq \mathbb{N}$ be any subset of the natural numbers, and let j be a positive integer. For each prime p_n, define the set

$$S = \{\, x \in \mathbb{N} : x \leq j \cap \text{the smallest prime factor of } x \text{ is } p_n \,\}.$$

S_j is the number of elements in the set $\{1, 2, \ldots, j\}$ that belong to S,

$$S_j = |\{x : x \in S \cap x \leq j\}|$$

The density of D_n, is

$$D_n = \lim_{j \to \infty} \frac{S_j}{j},$$

if this limit exists. We extend the definition of D_n from $[1, P_n]$ to the entire set of positive integers $[1, \infty)$.

1.7 Arithmetic Progression for the Limit as $j \to \infty$

Want to show:

$$D_n = \lim_{j \to \infty} \frac{S_j}{j} \;=\; \frac{\phi\big(P_{n-1}\big)}{P_n},$$

To have p_n be the smallest prime factor of an integer x, x must satisfy $x = p_n \cdot m$ where m shares no prime factor with $p_1, \ldots, p_{n-1}$. Among the integers $1, 2, \ldots, P_{n-1}$, exactly $\phi(P_{n-1})$, m satisfy this condition.

$$M_{n-1} \;=\; \big\{\, 1 \leq m \leq P_{n-1} \cap \gcd\big(m, P_{n-1}\big) = 1 \big\},$$

so $|M_{n-1}| = \phi(P_{n-1})$. Any integer x with smallest prime factor p_n must be of the form

$$x \;=\; p_n \cdot m \quad \text{for some } m \in M_{n-1},$$

plus possible multiples offset by shifts of length P_n. If

$$m \in M_{n-1} \quad \text{and} \quad r \in \mathbb{N}_0,$$

then

$$P_n \cdot r + p_n \cdot m$$

also has p_n as its smallest prime factor, since shifting by P_n does not introduce any smaller prime factor. Each integer in S_j appears in some line/arithmetic progression

$$P_n \cdot r + p_n \cdot m \quad \text{where } m \in M_{n-1}, \ \ r \geq 0.$$

1.8 Bounding $\frac{S_j}{j}$

Let j be large. We want to estimate how many integers $x \leq j$ belong to S_j. Let $q = \lfloor j/P_n \rfloor$. Then

$$q \cdot P_n \ \leq \ j \ < \ (q+1) \cdot P_n.$$

For each $m \in M_{n-1}$, the values of r range from $r = 0$ up to $r = q$.

$$P_n \cdot r + p_n \cdot m \ \leq \ j$$

We get at least $q\,\phi(P_{n-1})$, x, when r runs from 0 to q, and at most $(q+1)\,\phi(P_{n-1})$, x, because the leftover segment can add at most one more instance per $m \in M_{n-1}$.

$$q\,\phi(P_{n-1}) \quad \leq \quad S_j \quad \leq \quad \big(q+1\big)\,\phi(P_{n-1}).$$

$$\frac{q\,\phi(P_{n-1})}{j} \quad \leq \quad \frac{S_j}{j} \quad \leq \quad \frac{(q+1)\,\phi(P_{n-1})}{j}.$$

$$D_n = \lim_{j\to\infty} \frac{S_j}{j} \approx \frac{\phi(P_{n-1})}{P_n}.$$

This proves that the limit $\frac{S_j}{j}$ exists as $j \to \infty$ and is approximately $\frac{\phi(P_{n-1})}{P_n}$. We have extended density of p_n as the smallest prime factor from the interval $[1, P_n]$ to $[1, \infty)$. The largest error in counting up to some j is at most $\phi(P_{n-1})$ for $\frac{S_j}{j}$.

1.9 Conclusion

Assume p_n , then $D_n = \frac{\phi(P_{n-1})}{P_n}$. This is useful in problem contexts where divisibilty with respect to certain primes are necessary . Existing methods only utilize simple heuristics to determine the relative density ie: $\frac{1}{p_n}$, which is limited in its usefulness depending on the problem setup.

2 Approximating the behavior of D_n asymtotically for large n

We start from

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

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

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

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

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

where $\gamma \approx .577$ 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

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

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

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

Given

$$D_n \approx \frac{e^{-\gamma}}{p_n \ln p_n},$$

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

$$D_n \approx \frac{e^{-\gamma}}{n\ln n\cdot\ln(n\ln n)}$$

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

For large n , D_n will grow as $\ln n$ as opposed to $\ln\ln n$, since $\ln n \gg \ln\ln n$ for large n .

$$D_n \approx \frac{e^{-\gamma}}{n(\ln n)^2}$$

$$\begin{aligned} D_{11} &\approx \frac{e^{-\gamma}}{11\left(\ln 11\right)^2} \\ &\approx \frac{0.561459}{63.265169} \\ &\approx 0.008878 \end{aligned}$$

The limit of D_n as $j \to \infty$ is :

$$D_n \lim_{j\to\infty} = \lim_{j\to\infty}\frac{S_j}{j} \approx \frac{\phi\left(P_{n-1}\right)}{P_n} \approx \frac{e^{-\gamma}}{n(\ln n)^2} \approx \frac{e^{-\gamma}}{\infty} = 0$$

2.1 Conclusion

Assume $D_n = \frac{\phi(P_{n-1})}{P_n}$, then $D_n \approx \frac{e^{-\gamma}}{n(\ln n)^2}$. This is useful so that sums of D_n , may be approximated in terms of some large n or x by integration.

3 Sum of D_n

To determine the likelihood that an integer q is composite , we must consider the likelihood that q is divisible by any smaller prime. This can be done using a sum with respect to the densities of all smaller primes . The total sum of densities B_n of integers that are divisible by some prime p_i up to p_n, for which p_i is the smallest prime factor, is given by:

$$B_n = \sum_{i=1}^{n} \frac{\phi(P_{i-1})}{P_{i-1} \cdot p_i} = \sum_{i=1}^{n} \frac{\phi(P_{i-1})}{P_i} = \sum_{i=1}^{n} D_i \approx \sum_{i=1}^{n} \frac{e^{-\gamma}}{i(\ln i)^2}.$$

B_n is the measure of the likelihood of an integer q to share a factor with a prime less than or equal to p_n . Let $G = \{p_i \leq p_n | p_i\}$, then the likelihood that an integer is divisble by G is B_n . B_n strictly increases for subsequent n , because all terms are positive .

3.1 Divisibility by primes not in G

Let $K = \{p_i > p_n | p_i\}$. If q is not divisble by an element in G, then it must be divisble by an element in K , since G and K inlcude all primes, and via the fundamental theorem of arithmetic at least one of these primes must factor q. The likelihood that an integer is not factorable by G is equal to the tail sum of B_n , we will call this tail/end sum E_n . E_n is the likelihood that an element of K divides some q. In practice K is not the entire tail, however it is significantly many terms.

4 Divisibilty Decay Theorem

Assume we have E_n, E_{n+1}, then $E_n > E_{n+1}$.

4.1 Quantifying the Tail Sum of B_n as E_n

We have:

$$E_n = \sum_{m>n} D_m,$$

To estimate E_n, consider the sum:

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

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

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

4.2 Simplify and Evaluate the Integral

$$u = \ln x$$

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

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

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

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

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

Thus:

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

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

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

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

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

If M is extremely large the dominant term is $1/\ln n$, ensuring it remains on the order of $1/\ln n$ for large n. As $n \to \infty$, $E_n \to 0$.

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

$$\begin{aligned} E_{15} &\approx \frac{e^{-\gamma}}{\ln 15} \\ &\approx \frac{0.561459}{2.708050} \\ &\approx 0.2073 \end{aligned}$$

$$
\begin{aligned}
E_{16} &\approx \frac{e^{-\gamma}}{\ln 16} \\
&\approx \frac{0.561459}{2.772589} \\
&\approx 0.2026
\end{aligned}
$$

$$E_{15} > E_{16}$$

4.3 Conclusion

If we have E_n, E_{n+1}, then $E_n > E_{n+1}$. This is useful because it allows us to see that if some q is not divisible by G then the likelihood that it is divisible by K goes to 0, for large n.

5 Adjacent Coprime Divisibility

Assume P_n is the the product of the first n primes, then $P_n \pm 1$ is only divisible by K.
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 elements in G . It can be shown that $P_n \pm 2$ is factorable by 2 , $P_n \pm 3$ is factorable by 3 ... $P_n \pm p_n$ is factorable by p_n. 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. There are primes $p_{n+1}, p_{n+2}, ...$ in the interval $[p_n, P_n + 1]$, which may be a factor of $P_n \pm 1$. This is defined as $E_n = \sum_n^M F_m$, where $M = P_n + 1$. Let $K = \{P_n + 1 > p_m > p_n | p_m\}$ which is also the set of primes accounted for in E_n.
Assume P_n is the the product of the first n primes, then $P_n \pm 1$ is only divisible by K. This is useful because it constrains the likelihood that $P_n \pm 1$ is prime or composite, relative to only a single set of primes K, whose behavior of D_n and E_n for large n and M is already known.
If $P_n \pm 1$ is coprime to G and coprime to K then $P_n \pm 1$ is prime.

6 Quantifying the likelihood that either $P_n - 1$ or $P_n + 1$ are prime for subsequent n

Assume we have $P_n \pm 1, P_{n+1} \pm 1$, then the likelihood that $P_n \pm 1$ is prime, is less than the likelihood that $P_{n+1} \pm 1$ is prime.
As n increases , the likelihood of $P_n \pm 1$ being divisible by K decreases via the Divisibility Decay Theorem . Given that E_n is the likelihood of K to divide some positive integer and $E_n > E_{n+1}$, the likelihood that K is composite to $P_n \pm 1$ decreases for large n. The likelihood that either $P_n + 1$ or $P_n - 1$ is composite is approximately E_n for each , so the combined likelihood that either is composite is at most $\approx E_n + E_n = 2 \cdot E_n$. Therefore, as n increases, the likelihood that either $P_n \pm 1$ are composite decreases as $2 \cdot E_n$. It follows that as $n \to \infty$, the likelihood that both $P_n - 1$ and $P_n + 1$ are both not composite

is at least $(1 - 2 \cdot E_n)$, which increases for large n . Thus, the likelihood that $P_n \pm 1$ are both prime increases for large n as $(1 - 2 \cdot E_n)$.

6.1 Conclusion

Assume we have $P_n \pm 1, P_{n+1} \pm 1$, then the likelihood that $P_n \pm 1$ is prime, is less than the likelihood that $P_{n+1} \pm 1$ is prime. This is useful in locating large prime numbers with a decreased error rate.