Prime Numbers Density Approximation

Edward C. Larson
eclars1071@gmail.com

Abstract

A novel derivation of the density of the distribution of prime numbers is presented, based on a simple frequentist analysis and the smallest scale at which a rigorous upper bound on the frequency holds. An approximating differential equation is derived. It is shown that in the asymptotic limit, the density of primes, $\pi(x)$, scales as $x/\ln x$, in accordance with the Prime Number Theorem (PNT). The approach bridges the gap between discrete number theory and continuous differential modeling, offering a mechanistic explanation for the observed thinning of prime density that mirrors the foundational results of classical analysis.

Let p_n denote the n'th prime number (excluding unity), *viz.*,

$$p_1 = 2, p_2 = 3, p_3 = 5, p_4 = 7, p_5 = 11, \dots \tag{1}$$

Upper Limit on Primes Count

Consider the integers in any interval of the form

$$[a, a + P(n))$$

in which $a \geq 0$ and $P(n)$ is the product of the first n primes, *viz.*,

$$P(n) \equiv \prod_{i=1}^{n} p_i \tag{2}$$

These can be grouped into 2^n rows, indexed by the n-tuple $(b_1, \dots, b_n)$, in which b_k, for $1 \leq k \leq n$, denotes the truth value (0 for false, 1 for true) of the proposition that given $m \in [a, a + P(n))$ is divisible by p_k. The number of integers in the row indexed by $(b_1, \dots, b_n)$ is

$$\mathcal{N}(b_1, \dots, b_n) = \prod_{i=1}^{n} (p_i - 1)^{1-b_i} \tag{3}$$

The first row, with $b_k = 0 \ \forall k \in \{1, \dots, n\}$, is of special interest. It contains

$$Q(n) \equiv \mathcal{N}(0, \dots, 0) = \prod_{i=1}^{n} (p_i - 1) \tag{4}$$

elements. For $a > p_n$, all primes in $[a, a + P(n))$ are in the first row. $Q(n)$ is therefore a hard upper limit on the number of primes in the interval. It most closely approximates the actual number of primes in the interval for a slightly greater than p_n.

Estimation of Primes Density

We seek a continuous, differentiable function p, defined on $\{x \in \Re : x \geq 1\}$, such that $p(n)$ accurately approximates p_n for integer values of x.

$$\frac{1}{p'(n)} \approx \frac{1}{p_{n+1} - p_n} \tag{5}$$

represents the density of primes near p_n. Based on the argumentation in the preceding section, it can be approximated as

$$\frac{Q(n)}{P(n)} = \prod_{i=1}^{n} \left(1 - \frac{1}{p_i}\right) \tag{6}$$

The density decreases as n increases, which is reflected in the accumulation of factors less than unity on the right-hand side of Eq. 6.

Combining Eqs. 5-6, it follows that

$$\ln p'(n) \approx b - \sum_{i=1}^{n} \ln \left(1 - \frac{1}{p_i}\right) \tag{7}$$

in which $b \approx -0.8$ is a bias term that can be estimated empirically (see Appendix). Differentiation of both sides of Eq. 7 yields

$$\frac{p''(n)}{p'(n)} = -\ln \left(1 - \frac{1}{p(n)}\right) \tag{8}$$

which is a self-contained differential equation. The initial conditions for $p(1)$ and $p'(1)$ may be chosen freely, such that the resulting solution optimally matches p_n for integer values of n.

Solution Analysis of the Differential Equation

Taylor series expansion of the right-hand side of Eq. 8, to first order, yields

$$\frac{p''(n)}{p'(n)} = \frac{1}{p(n)} + \ldots \tag{9}$$

which holds accurately at large n and $p(n)$. Integration of Eq. 9 yields

$$p'(n) = \ln p(n) + \ln C \tag{10}$$

in which C is a constant that depends on the initial conditions, *viz.*,

$$C = e^{p'(1)}/p(1) \tag{11}$$

Integration of Eq. 10 yields

$$\int_{p(1)}^{p(n)} \frac{dy}{\ln y + \ln C} = n - 1 \tag{12}$$

Evaluation of the definite integral on the left-hand side of Eq. 12 yields the first-order approximation of $p(n)$, with dependencies on the initial conditions for $p(1)$ and $p'(1)$.

The definite integral expression in Eq. 12 is a transformation of the *logarithmic integral function*, *viz.*,

$$\mathrm{li}(x) \equiv \int_2^x \frac{dt}{\ln t} \tag{13}$$

Eq. 12 then becomes

$$\mathrm{li}(p(n)) \approx n \tag{14}$$

Asympotic Behavior of $p(n)$

In the asymptotic limit of $n \to \infty$, the logarithmic integral function behaves as

$$\mathrm{li}(x) \approx \frac{x}{\ln x} \tag{15}$$

It follows from Eqs. 14-15 that

$$p(n) \approx n \ln p(n) \tag{16}$$

and hence

$$\pi(x) \approx \frac{x}{\ln x} \tag{17}$$

in which $\pi(x)$ is the density of primes at x. It is the functional inverse of p, *viz.*,

$$\pi(p(n)) = n \tag{18}$$

The result in Eq. 17 is the celebrated Prime Number Theorem (PNT) [1], which holds asymptotically in the limit of $n \to \infty$. It was historically first conjectured by Gauss [2] and later validated by Hadamard and Vallée-Poussin.

The asymptotic distribution of prime numbers has been the subject of extensive study, and the PNT remains the principal result. Historically, the PNT was arrived at through complex analysis and the study of the Riemann zeta function, often treating the distribution as a statistical outcome of sieve methods. In contrast, the present analysis demonstrates that this asymptotic behavior is the natural consequence of a self-contained, second-order differential equation describing prime density as a dynamical process. This approach bridges the gap between discrete number theory and continuous differential modeling, offering a mechanistic explanation for the observed thinning of prime density that mirrors the foundational results of classical analysis.

Appendix: Empirical Estimation of Bias Term

The two sides of Eq. 7 are in tension - and the equation is approximate - because the left-hand side depends on actual differences between successive primes, based on the definition in Eq. 5. The left-hand side is therefore noisy. It can be as large as 1/2, even at arbitrary large n if the conjecture that there are infinitely many twin primes holds true. The right-hand side of the equation, on the other hand, is guaranteed to increase monotonically with each new prime.

The following Python program enables the difference between the two sides to be examined statistically over a large sample of primes, up to a specified limit (`N_max`). It provides estimates of the mean and standard deviation of the bias distribution. The mean of the distribution, which equates to the bias, is about -0.8 for modestly large primes. The bias is a useful corrective to compare the two sides of the equation, but it is not strictly well-defined in the limit of arbitrarily large numbers.

Code Sample:

```
import math
import statistics as sts

L_prime = [2]
dens_th = 0
L_bias = [math.log(1/2)]

N_max = 20000

j = 2
while j < N_max:
  isPrime = True
  for n in L_prime:
    if j % n == 0:
      isPrime = False
      break

  if isPrime:
    dens_em = -math.log(1 / (j - L_prime[-1]))
    dens_th -= math.log(1 - 1 / L_prime[-1])
    L_bias.append(dens_em - dens_th)
    L_prime.append(j)

    s = f'{j:6d}: {dens_em:0.6f}, {dens_th:0.6f}'
    print(s)

  j += 1

print('-----')
s = f'N_max  : {N_max:10d}\n'
s += f'N_prime: {len(L_prime):10d}\n'
s += 'Bias statistics from first half of primes list:\n'
s += f'mean = {sts.mean(L_bias[:len(L_bias) >> 1]):.6f}, '
s += f'stdev = {sts.stdev(L_bias[:len(L_bias) >> 1]):.6f}\n')
s += 'Bias statistics from second half of primes list:\n')
s += f'mean = {sts.mean(L_bias[len(L_bias) >> 1:]):.6f}, '
s += f'stdev = {sts.stdev(L_bias[len(L_bias) >> 1:]):.6f}\n')
print(s)
```

References

[1] Tom M. Apostol. *Introduction to Analytic Number Theory*. Provides rigorous foundational analysis of the Prime Number Theorem. Springer-Verlag, 1976.

[2] W. Gellert et al. *The VNR Concise Encyclopedia of Mathematics*. Van Nostrand Reinhold, 1977.