Prime number theorem

In this document, we are going to compare the performance of two well known formula approximating the exact count of primes less than a given integer n, namely, Gauss's and Legendre's formula which are given by:

- **Gauss** (also known as prime number theorem) [a] :

$$\pi(n) \sim \frac{n}{log(n)} \iff \lim_{n\to\infty} \frac{\pi(n)}{\left(\frac{n}{\log(n)}\right)} = 1$$

- **Legendre**:

$$\pi(n) \sim \frac{n}{log(n) - 1.08366}$$

where $\pi(n)$ is the prime-counting function (the number of primes less than or equal to n) and $\log(n)$ is the natural logarithm of n, $\ln(n)$. In fact **PNT** implies the following:

- For any constant b:

$$\pi(n) \sim \frac{n}{log(n) + b}$$

[a] Then it is was not a theorem since it was an educated guess without a proof. But it was proved later.

Before starting our discussion, I want to add two plots visualizing and "confirming" these approximations:

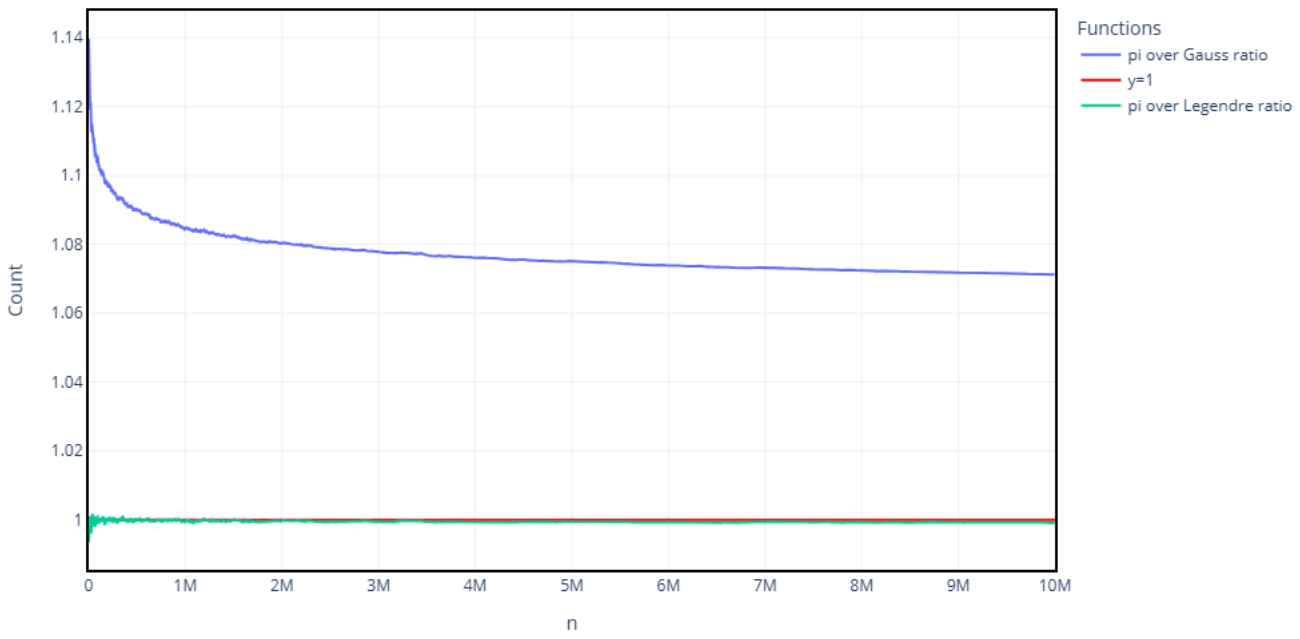

Figure 1: Ratios of $\pi(n)$ to Gauss and Legendre Approximations

Figure 1 represents: $\frac{\pi(n)}{\left(\frac{n}{\log(n)}\right)}$ and $\frac{\pi(n)}{\left(\frac{n}{\log(n)-1.08366}\right)}$ even though in this figure the ratio $\frac{\pi(n)}{\left(\frac{n}{\log(n)}\right)}$ seems to be far from 1, it is indeed tending to 1 for larger values and this is only because the convergence of this ratio is so slow compared to other approximations and due to lack of large values since they are computationally exhaustive, but the second ratio with Legendre corrective constant is clearly approaching 1 in a fast manner.

The second plot shows the prime counting function against both approximations for all values of n in range $[4, 2000]$:

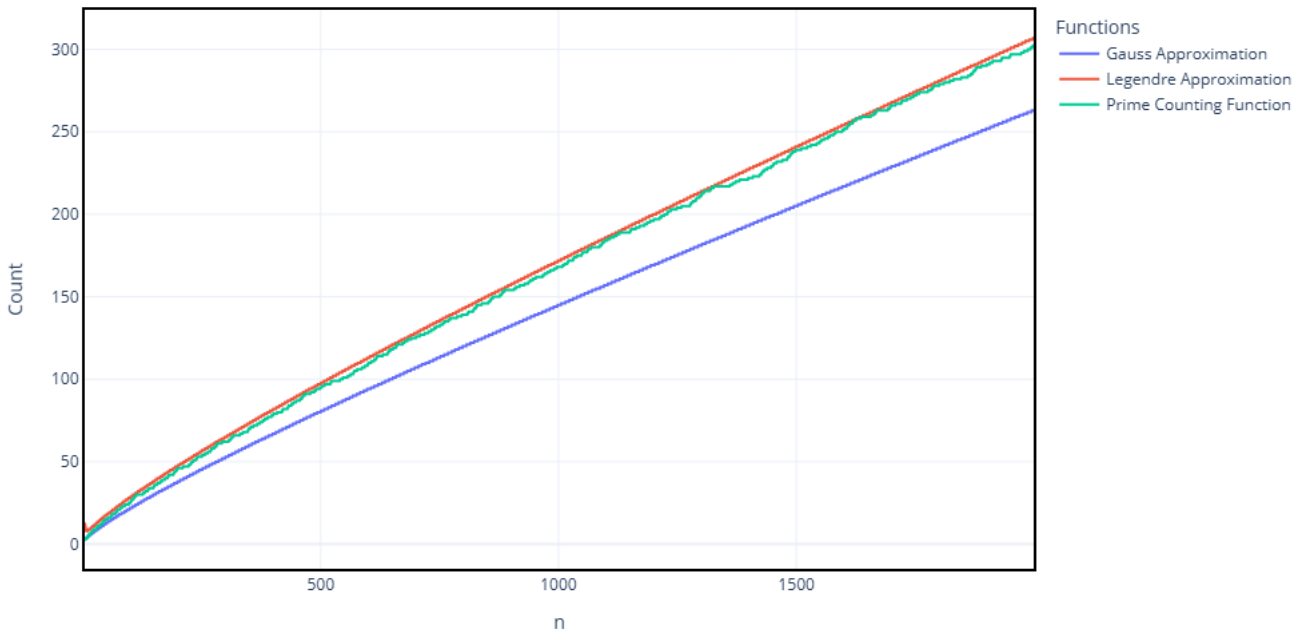

Figure 2: Prime counting approximations

Figure 2 shows that the function $\pi(n)$ follows a certain trend in the represented range and both formula quantifies this trend with a certain error. We also observe the role of Legendre regulating constant in giving better approximation.

These plots serve as validations of the previous statements but obviously, they does not represent any kind of proof.

The following table shows the exact count of primes alongside estimated values by both formula for values of n of the form $n = 10^k$ where k varies from 2 to 10:

n	10	10^2	10^3	10^4	10^5	10^6	10^7	10^8	10^9	10^{10}
$\pi(n)$	4	25	168	1229	9592	78498	664579	5761455	50847534	455052511
Gauss	4	22	145	1086	8686	72382	620421	5428681	48254942	434294482
Legendre	8	28	172	1231	9588	78543	665140	5768004	50917519	455743004

In order to compare these values we plot the absolute differences between approximations and actual prime counting function at the given points:

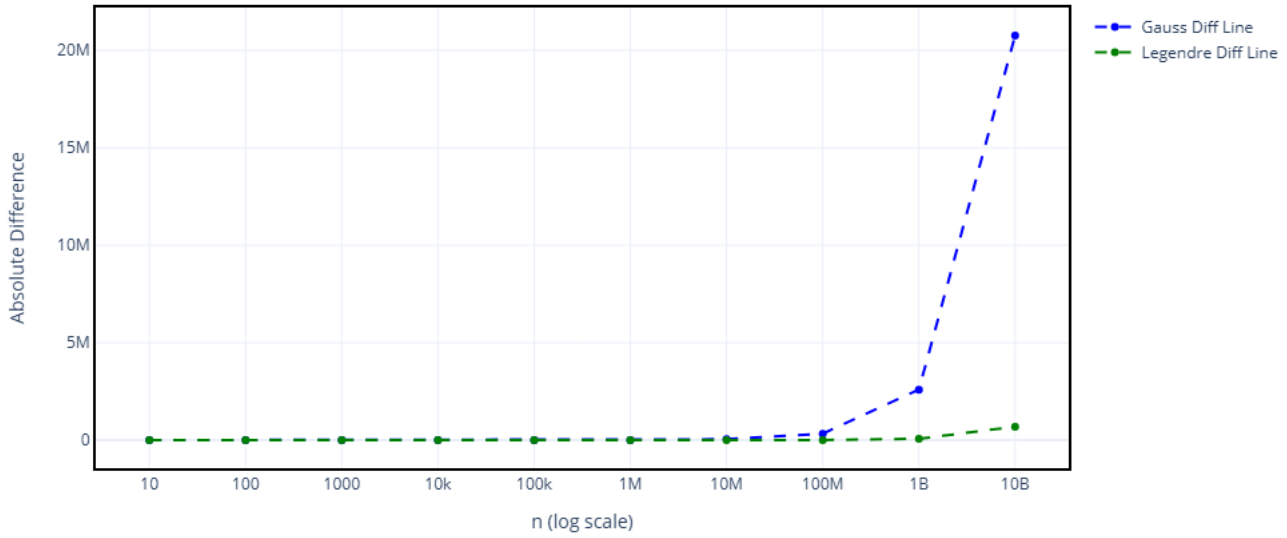

Figure 3: Absolute differences between approximations and $\pi(n)$

Figure 3 reveals that: For smaller values both formula provides good approximations but Gauss approximation is getting far from the exact values of $\pi(n)$ for very large numbers while Legendre approximation is keeping the gap relatively narrow even for larger values.

Important application

Here we will see an example showing how such approximations are useful to us:

Suppose we are using a RSA protocol and we want to generate a prime number with large bit-size d, then we will chose a random number of bit-size equals d and test if it is prime or not if yes we keep it if no we regenerate another random number, in general we are interested in how many numbers we expect to test before we find our prime number:

For simplicity, let $d = 1024$ and denote the random number by R then $R \in [2^{1023}, 2^{1024})$. **PNT** implies that the density of primes less than R is around $\frac{1}{\ln(2^{1024}-1)} \sim \frac{1}{\ln(2^{1024})} \approx \frac{1}{710}$. So we expect to test 710 number before we find one prime.

This only represents the basic idea because we know that prime numbers are not uniformly distributed among integers, in fact their density gets lower for larger values.

General remarks

- These approximations does not provide a bound of the error (the difference between $\pi(n)$ and both formula), as we have seen with Gauss formula previously.
- At the beginning, we mentioned that $\pi(n) \sim \frac{n}{log(n)+b}$ for any constant b(simple proof), it has been shown that $b = 1$ is the best choice in the long run. Also note that Legendre formula is particular case for $b = -1.08366$.

- **PNT** also implies that the n-th prime number is approximated by $n\ln(n)$(simple proof).

Conclusion

Based on what we have seen, Legendre formula offer better approximation of $\pi(n)$ then Gauss formul and both of them is interesting in our context (primality test) .