1 Distribition Of Primes

Given primes are so resistant to encoding into a simple generating formula, let's take a detour and try a different approach, **experimental mathematics**.

Number of Primes Up To A Number

We showed that primes don't run out as we explore larger and larger numbers. We also showed they don't thin out as quickly as the squares. So how quickly do they thin out?

One way to explore this is to keep a count of the number of primes as we progress along the whole numbers.

The expression $\pi(n)$ has become an abbreviation for 'the number of primes up to, and including, n'. For example, $\pi(5) = 3$ because there are 3 primes up to, and including, 5. The next number 6 is not prime, so $\pi(6)$ remains 3. The use of the symbol π can be confusing at first.

Figure 1.1 shows a chart of $\pi(n)$ for n up to 100. An fairly smooth curve seems to be emerging. This is slightly unexpected because the primes appear to be randomly placed amongst the numbers. The curve suggests the primes are governed by some kind of constraint. It wouldn't be too adventurous to say the curve looks logarithmic, like $\ln(n)$.
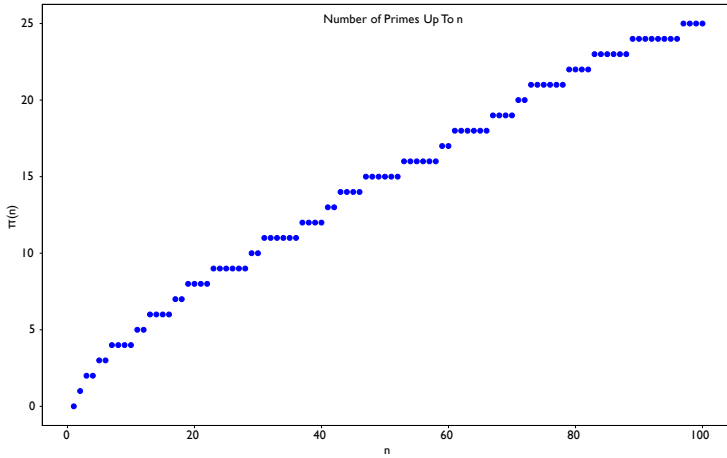
Figure 1.1: $\pi(n)$ for n from 1 to 100.

Rather Good Approximations for $\pi(n)$

Gauss, one the most prolific mathematicians in history, was the first to find an expression that approximates $\pi(n)$ fairly well. He was aged about 15 at the time.

$$\pi(n) \approx \frac{n}{\ln(n)}$$

The expression is suprisingly simple. It is worth pondering on what hidden pattern in the primes is captured by the natural logarithm $ln(n)$.

Just a year later, Gauss developed a different expression that approximates $\pi(n)$ even more closely.

$$\pi(n) \approx \int_0^n \frac{1}{ln(x)} dx$$

At first glance, this logarithmic integral function, shortened to li(n), appears to be a continuous form of the first approximation.

Figure 1.2 shows a comparison of these approximations with the actual $\pi(n)$ for n all the way up to 10,000. It's clear the logarithmic integral function is much closer to the actual prime counts.
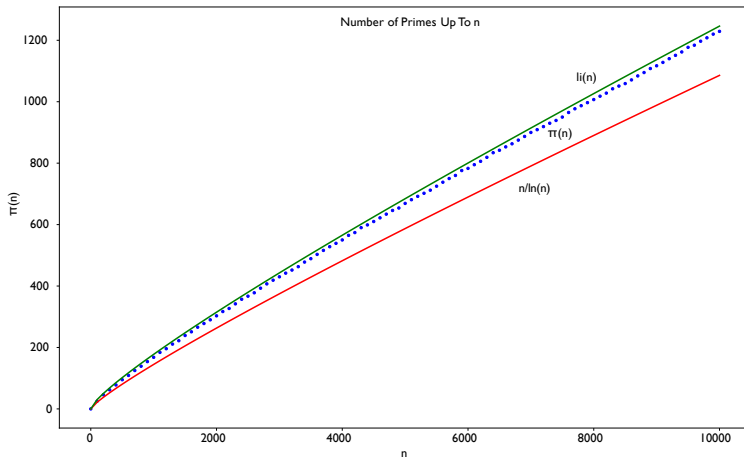

Figure 1.2: Comparing li(n) and $n/\ln(n)$ with $\pi(n)$.

Proportional Error

Looking again at the previous chart, the prime counting approximation $n/\ln(n)$ appears to be diverging away from the true prime count $\pi(n)$ as n gets larger. That is, the error appears to be getting ever larger. If we looked at the numbers, we'd also see li(n) diverging away from

$\pi(n)$ too. Does this mean the approximations become useless as n gets larger?

Figure 1.3 paints a different picture. It shows the error as a proportion of $\pi(n)$. We can see this proportional error becomes smaller as n grows to 10,000. It's also clear that $\mathrm{li}(n)$ has a distinctly smaller proportional error than $n/\ln(n)$.
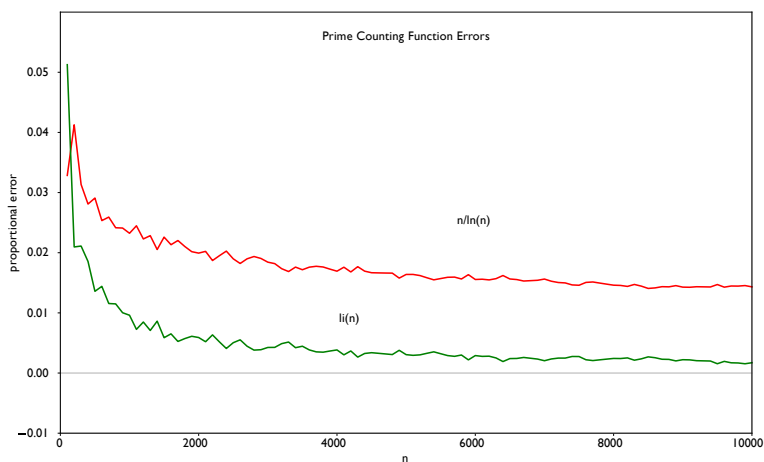

Figure 1.3: Proportional errors for $\mathrm{li}(n)$ and $n/\ln(n)$.

There are 1229 primes amongst the first 10,000 whole numbers. The logarithmic integral gives us $\mathrm{li}(10{,}000){=}1246$. The error is just 17, and as a proportion of 1229, an impressively small 0.0138.

If we extended n to even larger values, we'd find the proportional error would fall further towards zero. Perhaps these approximations are correct in the limit $n \to \infty$?

Prime Density

Let's look again at those approximations and see if we can interpret their form. The following compares Gauss' first approximation with a general expression for calculating the mass of a volume of stuff with a given average density.

$$\text{mass} = \text{density} \times \text{volume}$$

$$\pi(n) \approx \frac{1}{\ln(n)} \times n$$

The comparison suggests that $1/\ln(n)$ is the average density of primes. If true, this would be a remarkable insight into the primes.

We can apply a similar analogy to Gauss' second approximation too. This time we compare it with another general expression for calculating mass where the density is not assumed to be constant throughout its volume.

$$\text{mass} = \int (\text{density}) dv$$

$$\pi(n) \approx \int_0^n \frac{1}{ln(x)} dx$$

Again, $1/\ln(x)$ emerges as a more locally accurate density of primes around a number x.

It was this density of primes around a number that the young Gauss first noticed as he studied the number of primes in successive ranges of whole numbers, 1-1000, 1001-2000, 2001-3000, and so on.

Nth Prime?

If the density of primes is $1/\ln(n)$ then we can say average distance between primes is $\ln(n)$. We can then make the short leap to say the nth prime is approximately $n\ln(n)$.

Before we get too excited about having found a simple function for generating the nth prime, this expression is an approximation, based on an average, itself based on Gauss' approximation for $\pi(n)$.

It's still interesting to see how well this expression for the nth prime performs. Figure 1.4 shows the error in $n\ln(n)$ as a proportion of the actual n^{th} prime. As n increases to 10,000,000, the proportional error falls to a non-zero constant approximately 0.1017.
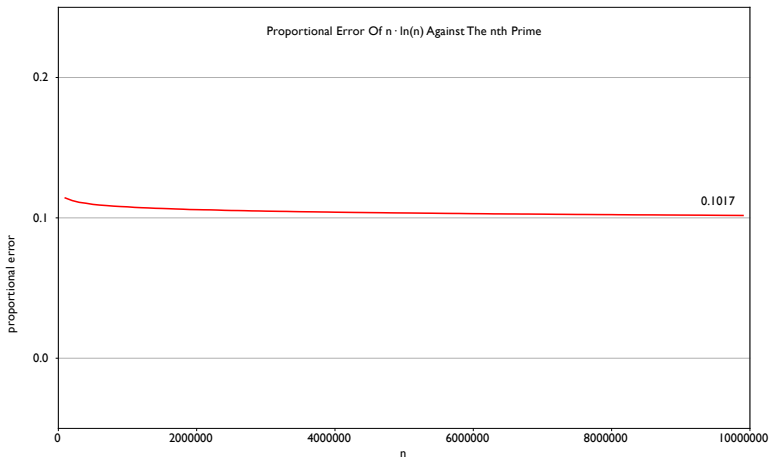

Figure 1.4: Proportional error of $n\ln(n)$ against the nth prime.

It's tempting to conclude the proportional error in $n\ln(n)$ approaches 0.1 as $n \to \infty$. We should be cautious because the error could be falling to zero very very slowly.

Anecodotal evidence is not mathematical proof, no matter how compelling it looks.

For example, our experiments show li(n) is always a bit higher than the true $\pi(n)$. If we extended n from 10,000 to 10,000,000 we'd still find li(n) was always higher. But in 1914 Littlewood proved that li(n) can become lower than $\pi(n)$ infinitely many times. More recent proofs show this starts to happen somewhere between 10^{19} and a mind-blowingly large 10^{316}.

Imperfect History

The question of who first developed an approximation for $\pi(n)$ is not perfectly clear. Gauss didn't always publish his work, leaving us to reconstruct history from notes and letters.

In his 1797 book on number theory, Legendre first published a form $n/(A\ln(n) + B)$, which he updated in his 1808 second edition to $n/(\ln(n) - 1.08366)$.

However, in 1849 Gauss wrote a letter to atronomer, and former student, Encke telling him that he had, in '1792 or 1793', developed the logarithmic integral approximation, which he wrote as $\int \frac{dn}{\log n}$. His collected works also reveal that in 1791 he had written about the simpler approximation, $\frac{a}{la}$ as he wrote it.

Appendix **??** presents reproductions of the relevant parts of these historical works.