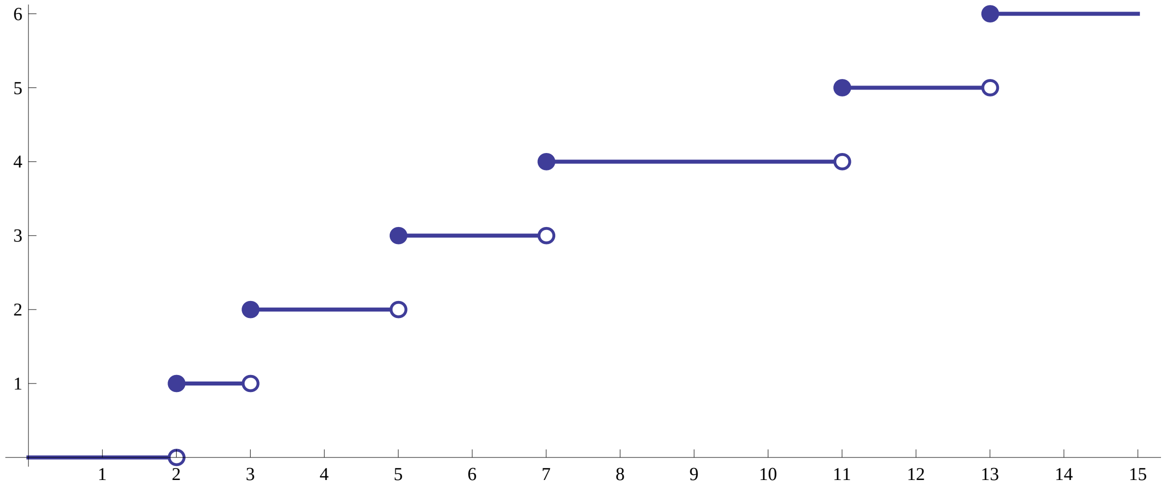
The Prime Number Theorem

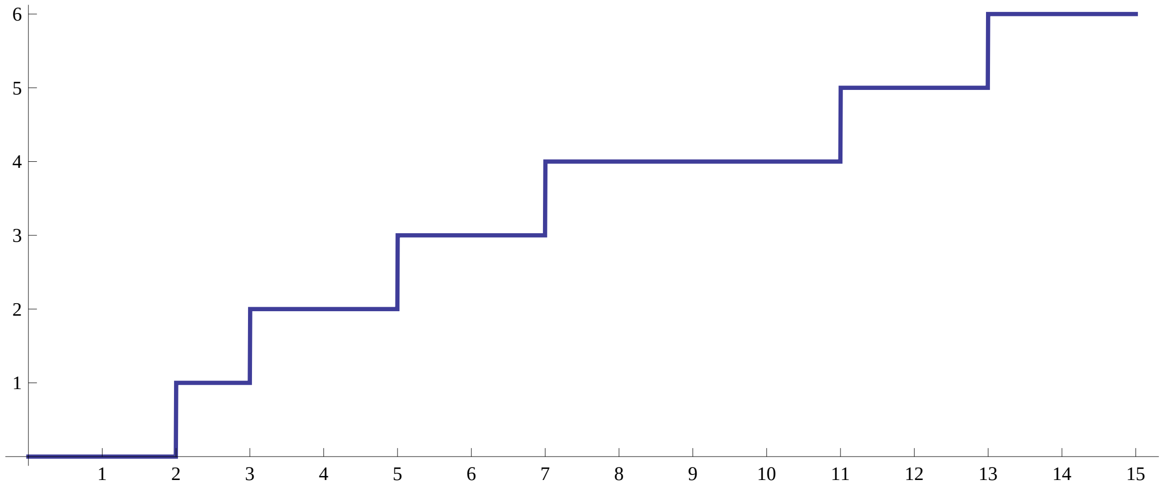
We study the distribution of primes via the function

$$\pi(x) = \text{the number of primes} \leq x$$



It's easier to draw this way:

$$\pi(x) = \text{the number of primes} \leq x$$

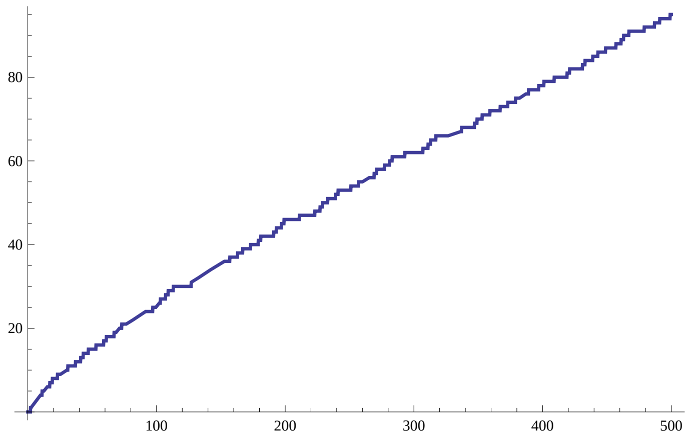


Clearly, $\pi(x) < x$

(not every number is prime)

and $\lim_{x \to \infty} \pi(x) = \infty$

(there are infinitely many primes)



How many primes are there among the first 1 000 000 000 integers?

By brute force, $\pi(1\,000\,000\,000) = 50\,847\,534$.

Can we approximate $\pi(x)$ using elementary functions?

Gauss's Prime Number Conjecture (1792)

$$\pi(x) \approx \frac{x}{\log(x)} \quad \text{for large } x.$$

x	$\pi(x)$	$\frac{x}{\log(x)}$	$\pi(x) - \frac{x}{\log(x)}$	$\frac{\pi(x) - x/\log(x)}{\pi(x)}$
10^2	25	21.7	3.3	0.1320
10^3	168	144.8	23.2	0.1381
10^4	1229	1085.7	143.3	0.1166
10^5	9592	8685.9	906.1	0.0945
10^6	78 498	72 382.4	6115.6	0.0779
10^7	664 579	620 421.0	44 158.3	0.0664

In terms of limits, we may write Gauss's 1792 conjecture as

$$\lim_{x\to\infty} \frac{\pi(x) - x/\log(x)}{\pi(x)} = 0 \qquad \text{or} \qquad \lim_{x\to\infty} \frac{x/\log(x)}{\pi(x)} = 1.$$

Gauss's Conjecture of 1849

For large x, $\quad \pi(x) \approx \int_2^x \frac{dt}{\log t}$

or

$$\lim_{x\to\infty} \frac{\int_2^x \frac{dt}{\log t}}{\pi(x)} = 1.$$

Aside: $\mathrm{Li}(x) \stackrel{\mathrm{def}}{=} \int_\mu^x \frac{dt}{\log t}.$

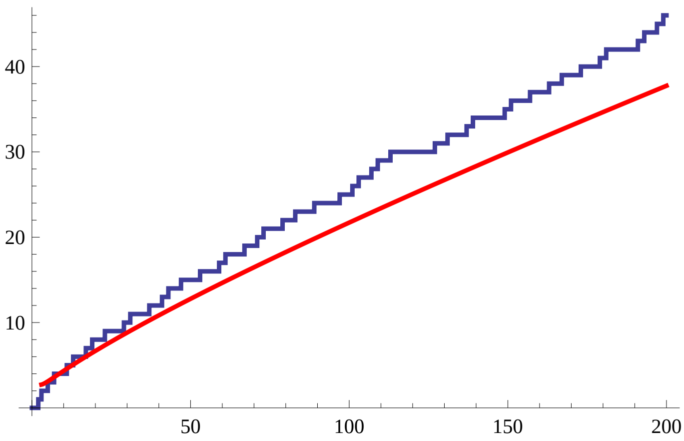
Exercise: $\lim_{x\to\infty} \dfrac{\left(\dfrac{x}{\log(x)}\right)}{\displaystyle\int_2^x \frac{dt}{\log t}} = 1.$ (Hint: l'Hôpital)

Corollary: If either $\lim_{x\to\infty} \dfrac{\left(\dfrac{x}{\log(x)}\right)}{\pi(x)}$ or $\lim_{x\to\infty} \dfrac{\displaystyle\int_2^x \frac{dt}{\log(t)}}{\pi(x)}$ exists, then

$$\lim_{x\to\infty} \frac{\left(\dfrac{x}{\log(x)}\right)}{\pi(x)} = 1 \qquad \text{is equivalent to} \qquad \lim_{x\to\infty} \frac{\displaystyle\int_2^x \frac{dt}{\log(t)}}{\pi(x)} = 1.$$

The Prime Number Theorem

$$\lim_{x\to\infty} \frac{\pi(x)}{\left(\frac{x}{\log(x)}\right)} = 1.$$



The convergence is very slow:

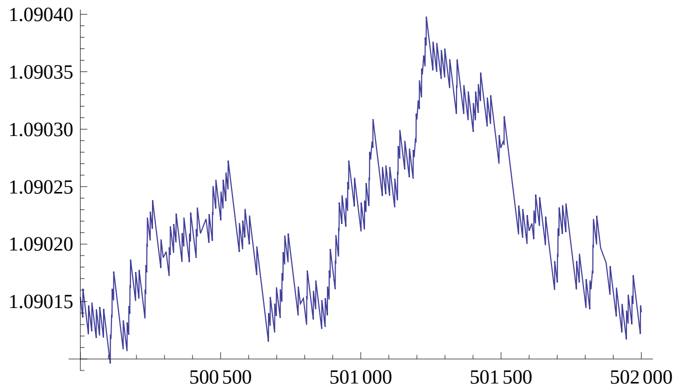
At $x = 10^6$, $\pi(x)/(x/\log(x)) \approx 1.08449$

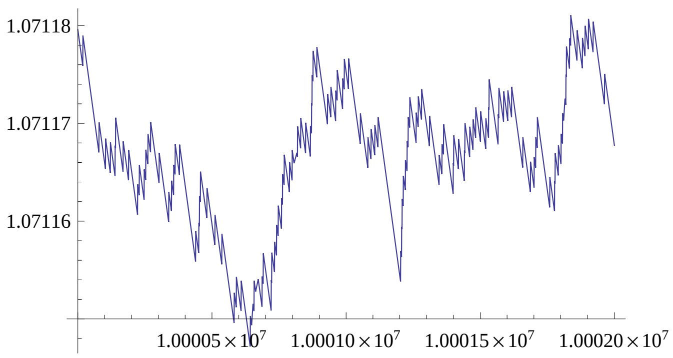
At $x = 10^9$, $\pi(x)/(x/\log(x)) \approx 1.05373$

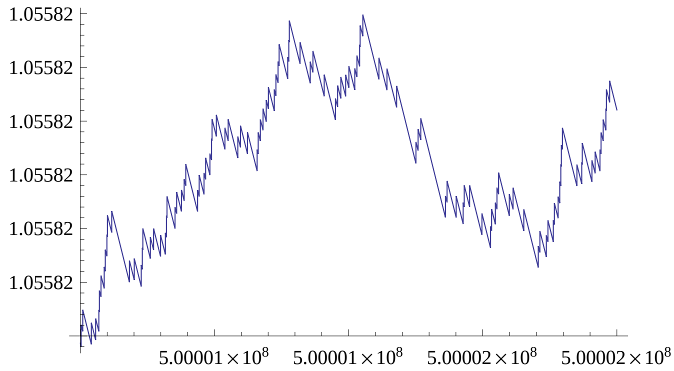
At $x = 10^{12}$, $\pi(x)/(x/\log(x)) \approx 1.03915$.

It's also messy.

$$y = \frac{\pi(x)}{\left(\frac{x}{\log(x)}\right)}$$







The first proofs of the Prime Number Theorem (1896) use the Riemann Zeta Function

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \frac{1}{1^s} + \frac{1}{2^s} + \frac{1}{3^s} + \cdots$$

Vallée-Poussin

From calculus, we know

$$\zeta(1) = \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots \qquad \text{diverges}$$

$$\zeta(2) = \frac{1}{1^2} + \frac{1}{2^2} + \frac{1}{3^2} + \cdots \qquad \text{converges to } \frac{\pi^2}{6}$$

$$\zeta(4) = \frac{1}{1^4} + \frac{1}{2^4} + \frac{1}{3^4} + \cdots \qquad \text{converges to } \frac{\pi^4}{90}$$

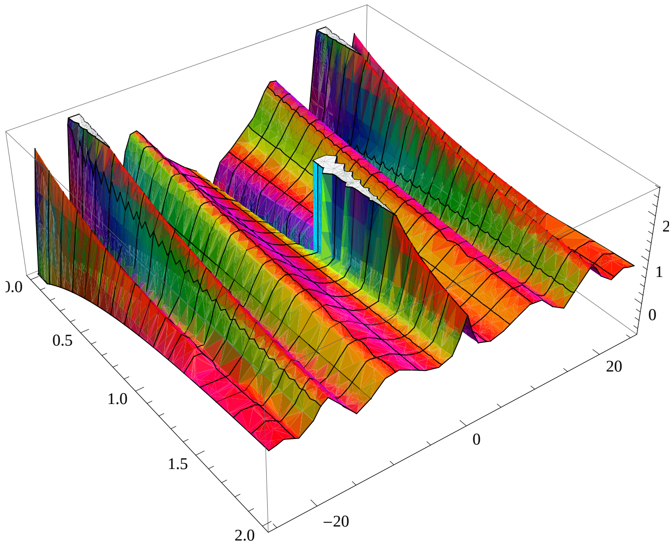
Hadamard

The function

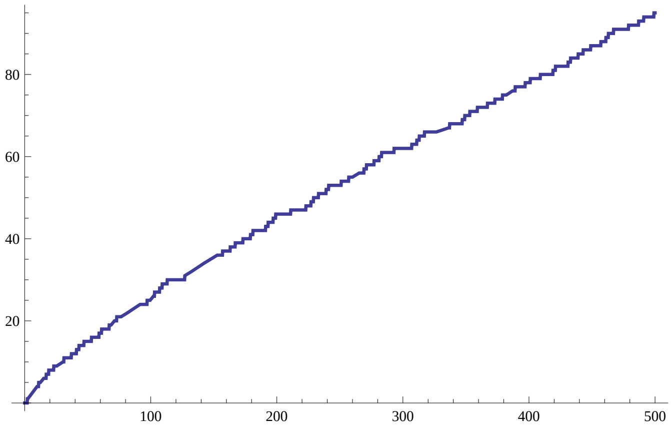
$\zeta(s)$ is defined for

all $s = x + iy$ in $\mathbb{C}$,

with a simple pole at $s = 1$.



What's the connection with primes?



Riemann

Fix a number M.

Let

$$\mathcal{P}_M = \{p_1, p_2, \ldots, p_r : p_i \le M\} = \text{set of primes} \le M.$$

Let

$$\Lambda_M = \{n \in \mathbb{Z}^+ : \text{prime factors of } n \text{ are in } \mathcal{P}_M\}.$$

Note that Λ_M is infinite and contains $\{1, 2, 3, \ldots, M\}$.

Example: $M = 8$.

Then

$$\mathcal{P}_8 = \{2, 3, 5, 7\}$$

and

$$\Lambda_8 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 15, 16, \ldots\} = \{2^{\alpha_1} 3^{\alpha_2} 5^{\alpha_3} 7^{\alpha_4} : \text{each } \alpha_i \text{ is a non-negative integer}\}$$

Recall that $\mathcal{P}_8 = \{2,\, 3,\, 5,\, 7\}$. Consider the product

$$\left(\frac{1}{2^0} + \frac{1}{2^1} + \frac{1}{2^2} + \cdots\right) \times \left(\frac{1}{3^0} + \frac{1}{3^1} + \frac{1}{3^2} + \cdots\right)$$
$$\times \left(\frac{1}{5^0} + \frac{1}{5^1} + \frac{1}{5^2} + \cdots\right) \times \left(\frac{1}{7^0} + \frac{1}{7^1} + \frac{1}{7^2} + \cdots\right)$$

Each term in the expansion has the form

$$\frac{1}{2^{\alpha_1}\, 3^{\alpha_2}\, 5^{\alpha_3}\, 7^{\alpha_4}} \quad \text{where} \quad \left(\begin{array}{l} \alpha_1,\, \alpha_2,\, \alpha_3,\, \alpha_4 \text{ are} \\ \text{are non-negative integers} \end{array} \right)$$

These denominators are exactly the elements of Λ_8, and we get

$$\prod_{p\,\in\mathcal{P}_8} \left(\frac{1}{p^0} + \frac{1}{p^1} + \frac{1}{p^2} + \cdots\right) = \sum_{n\in\Lambda_8} \frac{1}{n}.$$

Furthermore, by the geometric series theorem, we know that

$$\left(\frac{1}{p^0} + \frac{1}{p^1} + \frac{1}{p^2} + \cdots\right) = \sum_{r=0}^{\infty} \left(\frac{1}{p}\right)^r = \frac{1}{1 - \frac{1}{p}}.$$

So we get

$$\begin{aligned}
\sum_{n \in \Lambda_8} \frac{1}{n} &= \prod_{p \in \mathcal{P}_8} \left(\frac{1}{p^0} + \frac{1}{p^1} + \frac{1}{p^2} + \cdots\right) \\
&= \prod_{p \in \mathcal{P}_8} \frac{1}{1 - \frac{1}{p}} \\
&= \left(\frac{1}{1 - \frac{1}{2}}\right)\left(\frac{1}{1 - \frac{1}{3}}\right)\left(\frac{1}{1 - \frac{1}{5}}\right)\left(\frac{1}{1 - \frac{1}{7}}\right) = \frac{105}{24}.
\end{aligned}$$

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \frac{1}{9} + \frac{1}{10} + \frac{1}{12} + \frac{1}{14} + \frac{1}{15} + \frac{1}{16} + \frac{1}{18} + \frac{1}{20} + \frac{1}{21} + \cdots$$

More generally, we have

<u>Lemma</u>:

$$\prod_{p\in\mathcal{P}_M} \frac{1}{1-\frac{1}{p}} = \sum_{n\in\Lambda_M} \frac{1}{n}$$

<u>Extension of Lemma</u>:

$$\prod_{\text{all } p} \frac{1}{1-\frac{1}{p^s}} = \sum_{n=1}^{\infty} \frac{1}{n^s} = \zeta(s)$$

Now $\quad \Lambda_M \supset \{1,2,3,\ldots,M\}, \quad$ so $\displaystyle \sum_{n\in\Lambda_M} \frac{1}{n} > \sum_{n=1}^{M} \frac{1}{n}$.

This gives us the

<u>Corollary</u>:

$$\prod_{p\in\mathcal{P}_M} \frac{1}{1-\frac{1}{p}} > \sum_{n=1}^{M} \frac{1}{n}$$

As $M \to \infty$,

$$\prod_{p \in \mathcal{P}_M} \frac{1}{1-\frac{1}{p}} \to \prod_{\text{all primes}} \frac{1}{1-\frac{1}{p}} \qquad \text{and} \qquad \sum_{n=1}^{M} \frac{1}{n} \to \infty$$

so that $\displaystyle\prod_{\text{all primes}} \frac{1}{1-\frac{1}{p}} \to \infty,$

and we get

Corollary:

$$\prod_{\text{all } p} \left(1 - \frac{1}{p}\right) = 0$$

That is,

$$\frac{1}{2} \times \frac{2}{3} \times \frac{4}{5} \times \frac{6}{7} \times \frac{10}{11} \times \frac{12}{13} \times \frac{16}{17} \times \frac{18}{19} \times \frac{22}{23} \times \frac{28}{29} \times \cdots$$

can be made arbitrarily small.

A small digression:

Claim:

$$\prod_{\text{all primes}} \left(1 - \frac{1}{p}\right) = 0 \qquad \text{implies} \qquad \sum_{p \text{ prime}} \frac{1}{p} \text{ is divergent}$$

Remark: This gives another proof of the fact that there are infinitely many primes. The proof of the claim itself follows from a

Lemma: Let $1 > a_1 > a_2 > a_3 > \cdots$ be a decreasing sequence of positive numbers such that $\sum a_n$ converges. Then

$$\prod_{n=1}^{\infty}(1 - a_n) \neq 0$$

.

Proof sketch

Because $\sum a_n$ converges, we can find N so that $\displaystyle\sum_{n=N}^{\infty} a_n < \frac{1}{2}$.

$$S = \sum_{n=1}^{\infty} a_n = \underbrace{a_1 + a_2 + \cdots + a_{N-1}}_{> S - \frac{1}{2}} \; + \; \underbrace{a_N + a_{N+1} + \cdots}_{< \frac{1}{2}}$$

Write

$$\prod_{n=1}^{\infty}(1 - a_n) = \prod_{n=1}^{N-1}(1 - a_n) \times \prod_{n=N}^{\infty}(1 - a_n).$$

It will suffice to show that the second factor is non-zero.

Write

$$
\begin{aligned}
\prod_{n=N}^{\infty}(1-a_n) &= (1-a_N)(1-a_{N+1})(1-a_{N+2})\cdots \\
&= 1 \\
&\quad - \sum_i a_i \\
&\quad\quad + \sum_{i<j} a_i a_j \\
&\quad\quad\quad - \sum_{i<j<k} a_i a_j a_k \\
&\quad\quad\quad\quad + \cdots
\end{aligned}
$$

where all the indices start at N.

We can then show that

$$\sum_i a_i < \frac{1}{2}, \quad \sum_{i<j} a_i a_j < \frac{1}{2^2}, \quad \sum_{i<j<k} a_i a_j a_k < \frac{1}{2^3}, \quad \text{and so on,}$$

$$\text{and} \quad \sum_i a_i > \sum_{i<j} a_i a_j > \sum_{i<j<k} a_i a_j a_k > \cdots$$

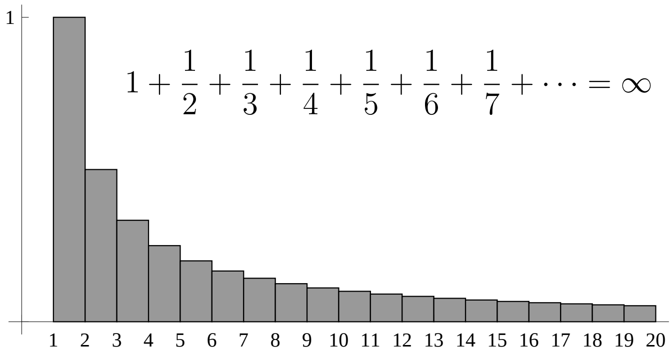
That is, the right side of

$$\prod_{n=N}^{\infty} (1 - a_n) = 1 - \sum_i a_i + \sum_{i<j} a_i a_j - \sum_{i<j<k} a_i a_j a_k + \cdots$$

satisfies the hypotheses of the Alternating Series Test.

By the AST, the series (and therefore the product) converges to a number between 1 and $1 - \sum_i a_i$. Since $\sum_i a_i < \frac{1}{2}$, we know that this number is non-zero. ■

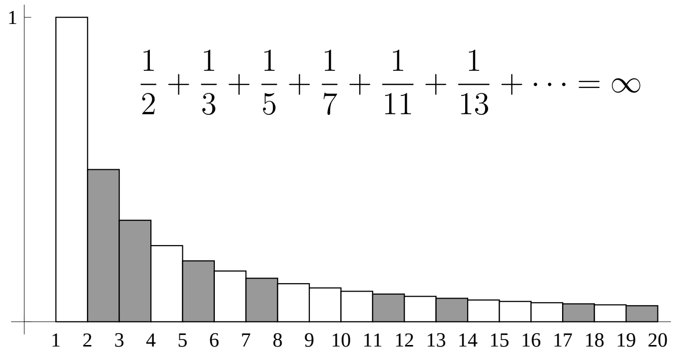
Old Stuff: $\sum \frac{1}{n}$ is divergent.



The harmonic series diverges *slowly*:

It takes about a million terms to get to a sum of 12.

New Stuff: $\sum \frac{1}{p}$ is divergent.

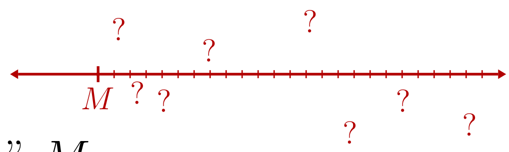


The $1/p$ series diverges *very slowly*:

A million terms gets you only to about 3.07.

But with either series, if you add up enough terms, you can make the sum as large as you want.

Estimating $\pi(x)$ Using Probability



Pick a random X with $X > M$ and X “near” M.

What’s the probability that you get a prime?

$$\Pr(X \text{ is not divisible by } 2) = \tfrac{1}{2} = \left(1 - \tfrac{1}{2}\right)$$
$$\Pr(X \text{ is not divisible by } 3) = \tfrac{2}{3} = \left(1 - \tfrac{1}{3}\right)$$
$$\Pr(X \text{ is not divisible by } 5) = \tfrac{4}{5} = \left(1 - \tfrac{1}{5}\right)$$
$$\vdots$$
$$\Pr(X \text{ is not divisible by any prime in } \mathcal{P}_M)$$

$$= \prod_{p \in \mathcal{P}_M} \left(1 - \frac{1}{p}\right) = \frac{1}{\displaystyle\sum_{n \in \Lambda_M} \frac{1}{n}}.$$

Heuristic 1:

Among the numbers between M and $M + \Delta M$, the proportion of primes is about

$$\frac{1}{\displaystyle\sum_{n\in\Lambda_M} \frac{1}{n}}.$$

There are approximately

$$\frac{\Delta M}{\displaystyle\sum_{n\in\Lambda_M} \frac{1}{n}}$$

primes in the interval

$$[M,\ M+\Delta M].$$

Example

$$\frac{100}{\displaystyle\sum_{n\in\Lambda_{5\,000\,000}} \frac{1}{n}} \approx 3.6398$$

and

$$\pi(5\,000\,100) - \pi(5\,000\,000) = 4.$$

Heuristic 2:

We can approximate $\displaystyle\sum_{n\in\Lambda_M} \frac{1}{n}$ with $\displaystyle\sum_{n=1}^{M} \frac{1}{n}$.

In fact,

$$\sum_{n\in\Lambda_M} \frac{1}{n} = \sum_{n=1}^{M} \frac{1}{n} + \sum_{\substack{n\in\Lambda_M \\ n>M}} \frac{1}{n}.$$

What happens to the second summation on the right as $M \to \infty$?

The number $\displaystyle\sum_{n=1}^{M} \frac{1}{n}$ is called the M^{th} *harmonic number*.

Heuristic 1 + Heuristic 2:

There are approximately

$$\frac{\Delta M}{\displaystyle\sum_{n=1}^{M} \frac{1}{n}} \quad \text{primes}$$

in the interval

$$[M,\ M + \Delta M].$$

Example

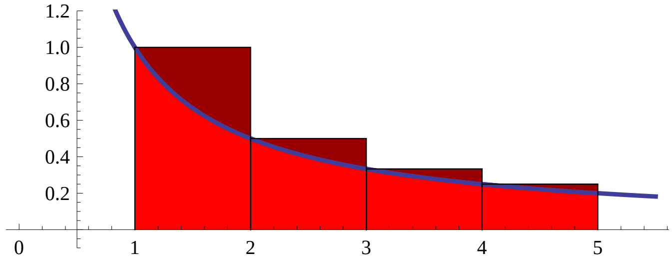
$$\frac{100}{\left(\displaystyle\sum_{n=1}^{2\,000\,000} \frac{1}{n}\right)} \approx 6.6287$$

and

$$\pi(2\,000\,100) - \pi(2\,000\,000) = 6.$$

Heuristic 3:

$$\sum_{n=1}^{M} \frac{1}{n} \approx \log(M+1).$$



The picture shows that $\sum_{n=1}^{4} \frac{1}{n}$ is equal to $\int_1^5 \frac{1}{x}\,dx$, which is $\log(5)$, plus the dark red "steps," whose total area is less than 1.

In fact, we have

$$\lim_{M\to\infty} \left[\sum_{n=1}^{M} \frac{1}{n} - \log(M+1) \right] \to \gamma$$

where $\gamma \approx 0.5772$.

So for large M,

$$\sum_{n=1}^{M} \frac{1}{n} \approx \log(M+1) + \gamma.$$

Heuristic 1 + Heuristic 2 + Heuristic 3:

There are approximately $\dfrac{\Delta M}{\log(M+1)+\gamma}$ primes in the interval $[M,\ M+\Delta M]$.

That is,

$$\begin{aligned} \pi(M+\Delta M)-\pi(M) &\approx \frac{\Delta M}{\log(M+1)+\gamma} \\ &\text{or} \\ \frac{\pi(M+\Delta M)-\pi(M)}{\Delta M} &\approx \frac{1}{\log(M+1)+\gamma} \\ &\text{or} \\ \frac{d}{dM}\left[\pi(M)\right] &\approx \frac{1}{\log(M+1)+\gamma}. \end{aligned}$$

Calculus?   In a Number Theory talk?

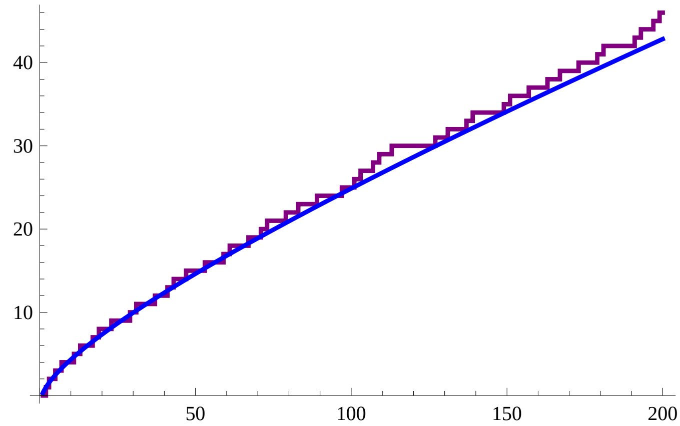
Now that we know (approximately) the derivative of $\pi(x)$, we can use the Fundamental Theorem of Calculus to find $\pi(x)$ itself.

Using the fact that $\pi(2) = 1$, we get

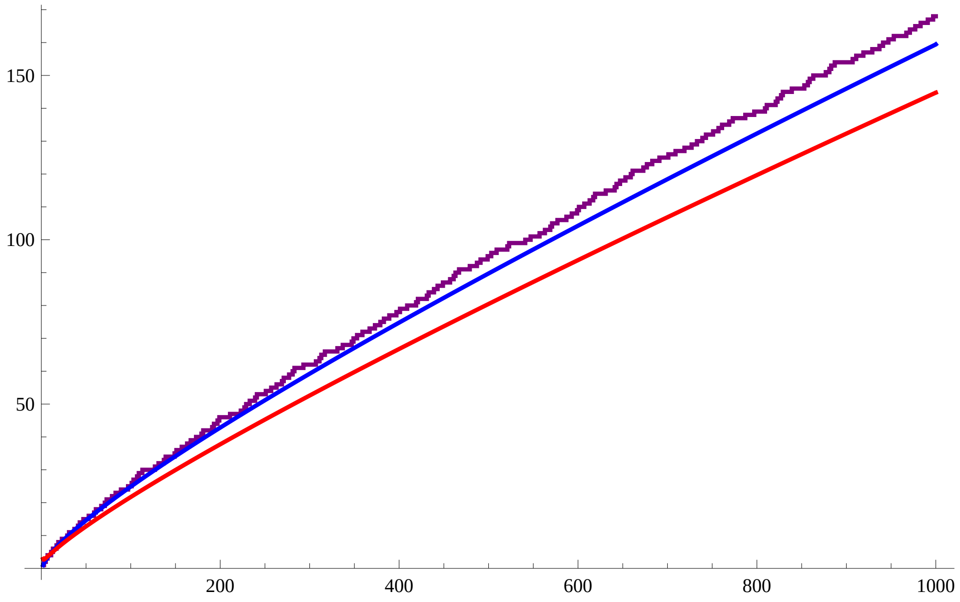
$$\pi(x) - \pi(2) \approx \int_2^x \frac{dt}{\log(t+1) + \gamma}$$

so that

$$\pi(x) \approx 1 + \int_2^x \frac{dt}{\log(t+1) + \gamma}.$$



Comparing the approximations



$\pi(x)$

$1 + \int_2^x \frac{1}{\log(t+1) + \gamma}\, dt$

$\frac{x}{\log(x)}$

Our approximation

$$\pi(x) \approx 1 + \int_2^x \frac{1}{\log(t+1)+\gamma}\, dt$$

can be "simplified" to

$$e^{-\gamma}\left[\mathrm{Ei}(\log(x+1)+\gamma) - \mathrm{Ei}(\log(3)+\gamma)\right],$$

where Ei is the exponential integral function,

$$\mathrm{Ei}(z) = -\int_{-z}^{\infty} \frac{e^{-t}}{t}\, dt.$$

But it can't be written in terms of elementary functions.

The good news is that you can use l'Hôpital's rule to show that

$$\lim_{x\to\infty} \frac{\left(1 + \int_2^x \frac{1}{\log(t+1)+\gamma}\,dt\right)}{\left(\frac{x}{\log(x)}\right)} \;=\; 1,$$

so by the real Prime Number Theorem,

our heuristic approximation also "works." That is,

$$\lim_{x\to\infty} \frac{\pi(x)}{\left(1 + \int_2^x \frac{1}{\log(t+1)+\gamma}\,dt\right)} \;=\; 1.$$

Some numbers

x	$\frac{x}{\log(x)}$	$1 + \int_2^x (\log(x+1) + \gamma)^{-1}\, dt$	$\pi(x)$
10^6	72 382	75 182	78 498
10^9	48 254 942	49 394 762	50 847 532
10^{12}	36 191 206 825	36 807 618 212	37 607 912 018
10^{15}	28 952 965 460 217	29 338 656 062 654	???