Prime Numbers Containing a Given String of Digits

An Application of the Prime Number Theorem

Dylan Nelson

Stellenbosch University Mathematics Society

8 April 2021

Outline

Reddit Post

- On 4 April 2016, a thread was posted to /r/math on reddit asking for the most surprising examples of divergent series.

Figure: `https://www.reddit.com/r/math/comments/4d879s/most_surprising_divergent_series/`

Reddit Post — Primes with a Prime Number of Digits

- In one example, we consider the set of prime numbers with a prime number of digits.
- It is claimed that the sum of the reciprocals of the elements in this set diverges.

Figure: https://www.reddit.com/r/math/comments/4d879s/most_surprising_divergent_series/d1oppgu

Reddit Post — Numbers Without a 9

- In another example, we consider all of the positive integers that *do not* have a 9 *anywhere* in their decimal expansion.
- In this case, it is claimed that the sum of the reciprocals of these numbers *converges*!

Figure: https://www.reddit.com/r/math/comments/4d879s/most_surprising_divergent_series/d1olh0o

Combining these Results

- I realised that a combination of (appropriate generalisations) of these two claims implies that there are infinitely many primes which have a prime number of digits, and which contain any given string of decimal digits that you like.
- And of course I promptly told everyone I know.
- I even wrote a blog post about it!

Figure: `http://www.mathemafrica.org/?p=12942`

This Talk

- In this talk, we will prove this result following the argument presented in the blog post.
- *BUT*... by considering convergent and divergent series, the blog post is needlessly circuitous.
- It is possible to give a more direct proof of a stronger result:

Proposition

Given a string of digits S, there is some natural number N, such that for all $n > N$, there is a prime with n digits that starts with S. (Or by some non-zero digit followed by S.)

- I will present a proof of this more general proposition towards the end of the talk.

Outline

The Harmonic Series

- One of the first somewhat surprising examples of a divergent series that students are shown is the *Harmonic Series*

$$\sum_{n=1}^{\infty} \frac{1}{n}.$$

- To show that this diverges, we group the terms in blocks of sizes equal to powers of 2, and then approximate each term by the smallest element in its block.

$$\begin{aligned}\sum_{n=1}^{\infty} \frac{1}{n} = 1 + \sum_{n=0}^{\infty} \sum_{k=2^n+1}^{2^{n+1}} \frac{1}{k} &\geq 1 + \sum_{n=0}^{\infty} \sum_{k=2^n+1}^{2^{n+1}} \frac{1}{2^{n+1}} \\ = 1 + \sum_{n=0}^{\infty} \frac{2^n}{2^{n+1}} &= 1 + \sum_{n=0}^{\infty} \frac{1}{2}.\end{aligned}$$

The Harmonic Series

- We of course get different behaviour if we sum the reciprocals of some subset of the natural numbers.
- The sum of the reciprocals of the powers of 2 converges:

$$\sum_{n=0}^{\infty} \frac{1}{2^n} = 2.$$

- The sum of the reciprocals of the prime numbers diverges.
- If we consider only the numbers that do not have a 9 in their decimal expansion, the sum of the reciprocals of these numbers *converges*.
- This feels surprising because it seems like there should be relatively few primes and many, many numbers without a 9 in their decimal expansion, but exactly the opposite is true.

Reciprocals of Numbers Without a Given String of Digits

- Let S be any string of digits. Let $\mathbb{N}_S$ be the set of natural numbers that contain S (contiguously) somewhere in their digits.
- We will show that

$$\sum_{n \notin \mathbb{N}_S} \frac{1}{n}$$

  converges.
- The approach will be similar to showing that the harmonic series diverges: we will group the digits in blocks of powers of $10^{\text{length of } S}$ and approximate the summands in each block by the largest element in the block.

Reciprocals of Numbers Without a Given String of Digits

- Let m be the length of S. We group together the numbers with between $km+1$ and $(k+1)m$ digits for some $k \geq 0$.

$$\sum_{n \notin \mathbb{N}_S} \frac{1}{n} = \sum_{k=0}^{\infty} \left(\sum_{\substack{10^{km} \leq n < 10^{km+1} \\ n \notin \mathbb{N}_S}} \frac{1}{n} \right) \leq \sum_{k=0}^{\infty} \left(\sum_{\substack{10^{km} \leq n < 10^{km+1} \\ n \notin \mathbb{N}_S}} \frac{1}{10^{km}} \right)$$

- To bound the size of this sum, we need an estimate for how many numbers with between $km+1$ and $(k+1)m$ digits do not contain S.

Estimating the Cardinality

- Consider a number n with between $km+1$ and $(k+1)m$ digits, and suppose that n does not contain S in its digits.
- Break the digits of n up into $k+1$ consecutive blocks of m digits. (One of the blocks may have fewer than m digits)
- There are 10^m possible blocks of m digits. Each block of digits of n can be any one of these possibilities *except* for S.
- There are thus at most

$$(10^m - 1)^{k+1}$$

possible values of n.

Bounding the Sum

- We see that

$$\sum_{n \notin \mathbb{N}_S} \frac{1}{n} \leq \sum_{k=0}^{\infty} \frac{(10^m - 1)^{k+1}}{10^{km}}$$

- since

$$\frac{10^m - 1}{10^m} < 1,$$

this is a geometric series, and converges!

$$\sum_{k=0}^{\infty} \frac{(10^m - 1)^{k+1}}{10^{km}} = (10^m - 1) \sum_{k=0}^{\infty} \left(\frac{10^m - 1}{10^m} \right)^k$$
$$= (10^m - 1) \frac{1}{1 - \frac{10^m - 1}{10^m}} = 10^m (10^m - 1)$$

Outline

The Prime Number Theorem

- In his talk on 1 April 2021, Lourens introduced the *Prime Number Theorem*:

Theorem (The Prime Number Theorem)

Let $\pi(x)$ denote the number of prime numbers that are less than or equal to the real number x. Then

$$\pi(x) \sim \frac{x}{\ln x}.$$

In other words,

$$\lim_{x\to\infty} \pi(x) \Big/ \frac{x}{\ln x} = 1.$$

The Prime Number Theorem

- Formally, this means that for every $\varepsilon > 0$, there exists $N > 0$ such that
$$(1-\varepsilon)\frac{x}{\ln x} \leq \pi(x) \leq (1+\varepsilon)\frac{x}{\ln x}$$
whenever $x > N$.
- One consequence of this is that if p_n is the n^{th} prime number, then $p_n \sim n \ln n$.
- Indeed, since $\pi(p_n) = n$, we have that
$$\lim_{n\to\infty} \frac{p_n}{n \ln n} = \lim_{n\to\infty} \frac{\ln p_n}{\ln n} \times \frac{p_n}{\ln p_n} \Big/ \pi(p_n).$$
- It is possible to show that $\lim_{n\to\infty} \ln p_n / \ln n = 1$, from which the result follows.

Proving that $\lim_{n\to\infty} \ln p_n / \ln n = 1$

- For all large enough x, we have that

$$\frac{1}{2}\frac{x}{\ln x} \leq \pi(x)$$

and so for large enough n we have that

$$p_n \leq 2n \ln p_n$$

which gives us that

$$\ln p_n \leq \ln 2 + \ln n + \ln(\ln p_n).$$

- It is thus enough to show that

$$\lim_{n\to\infty} \frac{\ln(\ln p_n)}{\ln n} = 0.$$

Proving that $\lim_{n\to\infty} \ln(\ln p_n)/\ln n = 0$

- Using our earlier estimate, we know that for large n,

$$\begin{aligned}\ln(\ln p_n) &\leq \ln\left(\ln 2 + \ln n + \ln(\ln p_n)\right)\\ &= \ln(\ln n) + \ln\left(\frac{\ln 2}{\ln n} + 1 + \frac{\ln(\ln p_n)}{\ln n}\right)\end{aligned}$$

and so it is enough to show that

$$\frac{\ln(\ln p_n)}{\ln n}$$

is bounded.

Proving that $\ln(\ln p_n)/\ln n$ is bounded

- If you ask me nicely, I'll prove that $p_n < 4^n$ for all n.
- It follows that

$$\ln(\ln p_n) < \ln\left(\ln\left(4^n\right)\right) = \ln\left(n \ln 4\right) = \ln n + \ln(\ln 4).$$

- Thus

$$\frac{\ln(\ln p_n)}{\ln n} < 1 + \frac{\ln(\ln 4)}{\ln n}$$

which is bounded.

The Sum of the Reciprocals of the Prime Numbers Diverges

- Consider the series

$$\sum_{p \text{ prime}} \frac{1}{p} = \sum_{n=1}^{\infty} \frac{1}{p_n}.$$

- If one is willing to use the Prime Number Theorem, then by the limit comparison test, this sum converges if and only if

$$\sum_{n=1}^{\infty} \frac{1}{n \ln n}$$

does.

The Sum of the Reciprocals of the Prime Numbers Diverges

- In turn, by the integral test, this sum converges if and only if the integral

$$\int_2^\infty \frac{1}{x \ln x}\,\mathrm{d}x$$

converges.

- But

$$\int_2^\infty \frac{1}{x \ln x}\,\mathrm{d}x = \ln(\ln x)\Big|_2^\infty \to \infty$$

and so the sum of the reciprocals of the prime numbers diverges.

A Generalisation

- We will show that the sum of the reciprocals of the primes with a prime number of digits diverges.
- In fact, we can prove an even more general result:

Proposition

Let S_0 be the set of natural numbers, and for each $n > 0$, let S_n be the set of prime numbers p where the number of digits in the decimal expansion of p is in S_{n-1}. Then

$$\sum_{p \in S_n} \frac{1}{p}$$

diverges.

- In particular, the fact that $\sum_{p \in S_2} 1/p$ diverges tells us that there are infinitely many prime numbers with a prime number of digits.

Proof

We prove the result by induction on n. For $n = 1$, the claim is that the harmonic series diverges, which we have already shown. Suppose that

$$\sum_{p \in S_n} \frac{1}{p}$$

diverges.

Proof

Let P_k be the set of prime numbers with k digits. Then

$$\sum_{p \in S_{n+1}} \frac{1}{p} = \sum_{k \in S_n} \sum_{p \in P_k} \frac{1}{p} \geq \sum_{k \in S_n} \frac{|P_k|}{10^k}.$$

It is thus sufficient to show that there is some constant C such that

$$\frac{|P_k|}{10^k} \geq \frac{C}{k}$$

for all large enough k.

Proof

By the Prime Number Theorem, there is some natural number N such that

$$\frac{1}{2}\frac{m}{\ln m} < \pi(m) < \frac{3}{2}\frac{m}{\ln m}$$

whenever $m > N$.
In particular, if $10^{k-1} > N$, then we have that

$$\pi\left(10^{k-1}\right) < \frac{3}{2}\frac{10^{k-1}}{(k-1)\ln 10}$$

and

$$\pi\left(10^{k}\right) > \frac{1}{2}\frac{10^{k}}{k\ln 10}.$$

Proof

For such a k, we have that

$$\begin{aligned}\frac{|P_k|}{10^k} &= \frac{\pi\left(10^k\right) - \pi\left(10^{k-1}\right)}{10^k} \\ &> \frac{1}{10^k}\left(\frac{1}{2}\frac{10^k}{k\ln 10} - \frac{3}{2}\frac{10^{k-1}}{(k-1)\ln 10}\right) \\ &= \frac{1}{20\ln 10}\frac{10(k-1)-3k}{k(k-1)} = \frac{1}{20\ln 10}\frac{7k-10}{k(k-1)}.\end{aligned}$$

For any constant $A < 7$, we have that $7k - 10 > A(k-1)$ provided that k is large enough, and then taking $C = \frac{1}{20\ln 10}$, we have that

$$\frac{|P_k|}{10^k} > \frac{C}{k}$$

for all large enough k.

Outline

Combining these Results

- Let S be some string of decimal digits.
- As before, S_n is the set of primes where the number of digits is prime, the number of digits of the number of digits is prime, and so on.
- Let $\mathbb{N}_S$ be the set of natural numbers that contain S somewhere in their decimal expansion.
- Suppose that $S_n \cap \mathbb{N}_S$ is finite. Then there is some natural number N such that if $p > N$ and $p \in S_n$, we have that $p \notin \mathbb{N}_S$.
- It follows that

$$\sum_{\substack{p>N \\ p\in S_n}} \frac{1}{p} \leq \sum_{\substack{p>N \\ p\notin \mathbb{N}_S}} \frac{1}{p}.$$

- But the sum on the left diverges, while the sum on the right converges. A contradiction!

Outline

A More General Result

- The fact that there are infinitely many prime numbers that have a prime number of digits and that contain your phone number somewhere among their digits is also a consequence of the following more general result.

Proposition

Given a string of digits S, there is some natural number N, such that for all $n > N$, there is a prime with n digits that starts with S. (Or by some non-zero digit followed by S.)

- This shows that having a prime number of digits isn't special. There is an appropriate prime number with almost every number of digits.

Proof

- Let m be the natural number whose decimal representation is S. If S starts with a 0, instead let m be the number whose decimal representation is 1 followed by S.
- We want to show that for all large enough natural numbers n, there is a prime between $10^k m$ and $10^{k+1} m - 1$.
- Equivalently, we want to show that

$$\pi\left(10^{n+1} m - 1\right) - \pi\left(10^n m - 1\right) > 0$$

  for all large enough n.
- Since $10^n m$ is never prime, this is the same as proving that

$$\pi\left(10^{n+1} m\right) - \pi\left(10^n m\right) > 0.$$

Proof

- For all large enough x, we know that
$$\frac{1}{2}\frac{x}{\ln x} < \pi(x) < \frac{3}{2}\frac{x}{\ln x}.$$
- It follows that for all large enough n, we have that
$$\pi\left(10^{n+1}m\right) - \pi\left(10^{n}m\right) > \frac{1}{2}\frac{10^{n+1}m}{\ln\left(10^{n+1}m\right)} - \frac{3}{2}\frac{10^{n}m}{\ln\left(10^{n}m\right)}.$$
- We wish to show that this is positive for large enough n. Since
$$\frac{10^{n}m}{2\ln\left(10^{n+1}m\right)\ln\left(10^{n}m\right)}$$
is positive, this is equivalent to showing that
$$10\ln\left(10^{n}m\right) - 3\ln\left(10^{n+1}m\right)$$
is positive for large n.

Proof

- We have that

$$\begin{aligned}
10 \ln \left(10^n m\right) - 3 \ln \left(10^{n+1} m\right) \\
&= 10 \left(n \ln 10 + \ln m\right) - 3 \left((n+1) \ln 10 + \ln m\right) \\
&= 7n \ln 10 + 7 \ln m - 3 \ln 10
\end{aligned}$$

  which is positive for all positive integers n.

- It follows that as long as n is large enough that $x = 10^n$ satisfies the bound

$$\frac{1}{2} \frac{x}{\ln x} < \pi(x) < \frac{3}{2} \frac{x}{\ln x},$$

  we have that there is a prime with $n + \text{length}(S)$ (possibly $+1$) digits that starts either with S, or with 1 followed by S.

Outline

Did We Actually Need the Prime Number Theorem?

- Other than when showing that the reciprocals of the primes diverges, the only consequence of the Prime Number Theorem that we have used is that for large x, we have that

$$\frac{1}{2}\frac{x}{\ln x} < \pi(x) < \frac{3}{2}\frac{x}{\ln x}.$$

- In fact, even weaker bounds would probably have been sufficient. We did not actually need the full power of the Prime Number Theorem.
- For some of these bounds, much more elementary proofs are known.
- Perhaps they can be explored in an upcoming talk *"It's Prime Time ∞: Elementary Proofs of Prime Number Theorem-like Results."*
- They also happen to show up in the Honours course in Analytic Number Theory.

Outline

Website

- I took some time to implement the Bailie-PSW pseudo-primality test in javascript.
- This allowed me to create a site where you can enter a string of digits, and it will find up to 100 prime numbers with a given number of digits containing the given string of digits.
- You can try it out here:

Figure: `https://dlnnlsn.github.io/prime-strings`