A Taste of Analytic Number Theory

Ayan Nath*

ayan.nmath

This article, aimed at olympiad contestants, focuses on solving olympiad number theory problems using analytic techniques and making contestants familiar with common techniques and results in this topic. We start with the prime number theorem, give an elementary proof of the weak version and establish a few well known estimates for the two Chebyshev functions. We also show Mertens' first theorem on the fly and discuss Mertens' second theorem. Asymptotic density, equidistribution theorem are also added.

§1 The Prime Number Theorem

§1.1 Introduction

Primes are the building blocks of the integers, just as molecules and atoms are the building blocks of nature, hence it makes great sense to study about primes, and in particular, distribution of primes. I think you know that there are infinitely many primes and most likely you already know Euclid's proof to it, but it doesn't tell us anything significant about the distribution of primes. This is exactly what our objective is, i.e. try to understand the distribution of primes. It is natural to define the function

$$\pi(x) = \text{No. of primes at most } x.$$

We would like to find a "formula" for $\pi(x)$ in terms of x, it turns out finding an exact formula is not really possible due to the raggedy nature of primes. Instead we try to estimate it. The well known estimate which we call the Prime number theorem (PNT) asserts that:

Theorem 1.1 Prime Number Theorem (PNT) —

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

here we write $f(x) \sim g(x)$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$, the way to think about this is, $f(x)$ is approximated by $g(x)$, the larger the x, the better the approximation. We do not prove this here, but instead we establish a weaker estimate that there exist positive real numbers a and b such that

$$\frac{ax}{\log x} < \pi(x) < \frac{bx}{\log x}.$$

Before we move on to the proof, it will help to get comfortable with big-$\mathcal{O}$ and little-o notation:

*AoPS user : https://artofproblemsolving.com/community/user/362567

Digression 1.2. If f and g are two functions then we say that $f(x)$ is $\mathcal{O}(g(x))$ if and only if there exist some constant C so that $|f(x)| < Cg(x)$ for all large x. And we say that $f(x)$ is $o(g(x))$ if and only if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$ where $g(x)$ should be non-zero for all large enough x. For example, $\sin x + e^{91}$ is $\mathcal{O}(1)$, $x + (\log x)^{10}$ is $\mathcal{O}(x)$, $\sin x + \log x$ is $o(x^{0.001})$. If $f(x)$ is $\mathcal{O}(g(x))$ then this is often expressed as $f(x) = \mathcal{O}(g(x))$ and similarly $f(x) = o(g(x))$ if $f(x)$ is $o(g(x))$, but you should remember that this is an abuse of notation. For example, we write $\lfloor x \rfloor = x + \mathcal{O}(1), \log x = o(x^{0.001}), n! = \mathcal{O}(n^n)$ in this article. It will help to get comfortable with these notations, it lets us to not compute stuff we don't care about. I suggest reading this : https://en.wikipedia.org/wiki/Big_O_notation

§1.2 Proving weak PNT

Let me define some functions, bear with me for now, I will explain the motivation in a moment. Define the **von Mangoldt function** $\Lambda : \mathbb{N} \to \mathbb{R}$ as

$$\Lambda(n) = \begin{cases} \log p, & n = p^k \text{ for some prime } p \text{ and positive integer } k \\ 0, & \text{otherwise} \end{cases}$$

one can also think of this as weighting all the prime powers p^k with $\log p$. You can easily see that

$$\sum_{d|n} \Lambda(d) = \log n.$$

Whenever we are trying to find bounds it is a common theme to look at the "big picture" at once, also called "global" methods, or in simple terms, double counting. Since we are looking for bounds related to primes, it is somewhat motivated to "sum" everything up and try to double count:

$$\begin{aligned} \sum_{n\le x} \log n &= \sum_{n\le x} \sum_{d|n} \Lambda(d) \\ &= \sum_{d\le x} \Lambda(d) \sum_{n\le x, d|n} 1 \\ &= \sum_{d\le x} \Lambda(d) \left\lfloor \frac{x}{d} \right\rfloor \\ &= \sum_{d\le x} \Lambda(d) \left(\frac{x}{d} + \mathcal{O}(1) \right) \\ &= x \sum_{n\le x} \frac{\Lambda(n)}{n} + \mathcal{O}(1) \sum_{n\le x} \Lambda(n) \end{aligned}$$

Now where did we double count? We double counted when we swapped the summations. The left hand side is very easy to estimate accurately (Unimportant: Those who know integration be like - lol just integrate). For now let us focus on the RHS, notice how the RHS is related to primes while the LHS is not, clearly the above equation has some information about primes encoded via the von Mangoldt function. We would want to estimate $\sum_{n\le x} \Lambda(n)$ to get rid of the awkward $\mathcal{O}(1)$ multiple. It now makes sense to define

$$\psi(x) = \sum_{n\le x} \Lambda(n).$$

This is called the **Second Chebyshev Function**, we take the domain as $\mathbb{R}$ instead of $\mathbb{N}$ to avoid writing floors whenever we have a non-integral input, we do this with almost all

the discussed functions in this article. In what follows, p always denotes a prime number. Observe the following *rough* calculation

$$\psi(x) = \sum_{n\leq x} \Lambda(n) = \sum_{p\leq x} \log p \left\lfloor \log_p x \right\rfloor = \sum_{p\leq x} \log p \left\lfloor \frac{\log x}{\log p} \right\rfloor \approx \sum_{p\leq x} \log x = \pi(x)\log x.$$

This suggests that $\psi(x) \sim x$ (which is indeed true). We are deliberately ambiguous about what $\approx$ means. We just need a rough estimate for $\psi(x)$, even $\psi(x) = \mathcal{O}(x)$ should do. For now assume this is true, we will get that

$$\sum_{n\leq x} \log n = x \sum_{n\leq x} \frac{\Lambda(n)}{n} + \mathcal{O}(x).$$

We would now want an estimate for the LHS.

Lemma 1.3 Weak Stirling's Approximation – $\sum_{n\leq x} \log n > x\log x - x$ for all $x \in \mathbb{N}$ and in particular, $\sum_{n\leq x} \log n = x\log x + \mathcal{O}(x)$ for all $x \in \mathbb{R}$.

Proof. The most common and natural way to prove this would be direct integration but we won't do that here. Look at the expansion of e^x where $x > 0$ is an integer:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

Clearly $\frac{x^x}{x!}$ is a term in the expansion, therefore

$$e^x > \frac{x^x}{x!} \implies x > x\log x - \log x! \implies \sum_{n\leq x} \log n > x\log x - x.$$

The lemma is proved now because $\sum_{n\leq x} \log n < x\log x$ is trivial. □

Using the above lemma we have that

$$x\log x + \mathcal{O}(x) = x\sum_{n\leq x} \frac{\Lambda(n)}{n} + \mathcal{O}(x) \implies \sum_{n\leq x} \frac{\Lambda(n)}{n} = \log x + \mathcal{O}(1).$$

The above relation is clearly useful since the LHS encodes primes in it, and gives us an estimate of a "global" sum involving primes. What is left is to prove that $\psi(x) = \mathcal{O}(x)$. The main idea is to write

$$\begin{aligned}\psi(x) &= \sum_{p\leq x} \log p + \sum_{p^2\leq x} \log p + \sum_{p^3\leq x} \log p + \cdots \\ &= \sum_{p\leq x} \log p + \sum_{p\leq \sqrt{x}} \log p + \sum_{p\leq \sqrt[3]{x}} \log p + \cdots \qquad (\star)\end{aligned}$$

For brevity let us define

$$\theta(x) = \sum_{p\leq x} \log p.$$

This is called the **First Chebyshev Function**, again the domain here is $\mathbb{R}$. Another way to motivate the first Chebyshev function is :

$$\begin{aligned}\psi(x) = \sum_{n\leq x} \Lambda(n) &= \sum_{p\leq x} \log p \left\lfloor \log_p x \right\rfloor \\ &= \sum_{p\leq x} \log p \left(\frac{\log x}{\log p} + \mathcal{O}(1)\right) \\ &= \pi(x)\log x + \mathcal{O}(1) \underbrace{\sum_{p\leq x} \log p}_{\theta(x)}.\end{aligned}$$

The above relation also suggests that $\theta(x) \sim x$ (which is indeed true, but we don't need that here). We can write $(\star)$ concisely as

$$\psi(x) = \theta(x) + \theta(x^{1/2}) + \theta(x^{1/3}) + \cdots$$

This sum is not infinite, the terms become zero eventually. Note that $\theta(x)$ is the "biggest" term in the RHS and the rest of them are "small", let us try to prove $\theta(x) = \mathcal{O}(x)$, the fact that $\psi(x) - \theta(x)$ is "small" (compared to $\mathcal{O}(x)$ of course) will be automatically implied.

Lemma 1.4 – $\theta(n) < 4n\log 2$, in particular $\theta(x) = \mathcal{O}(x)$.

Proof. The main idea is to consider a number which is divisible by many consecutive primes but the size of the number is not too large. A crude example is $x!$. Clearly the primes less than x divide $x!$, therefore $\prod_{p\leq x} p \leq x! \implies \theta(x) \leq \log x! = x\log x + \mathcal{O}(x)$, yes this is a stupidly trivial bound, but the point is that we want to do something similar.

We consider $\binom{2n}{n}$. Note that this number is divisible by all primes in the interval $[n+1, 2n]$. Therefore, $\prod_{n<p\leq 2n} p \leq \binom{2n}{n}$. By binomial theorem, $\binom{2n}{n}$ is trivially bounded above by $(1+1)^{2n} = 2^{2n}$. Therefore, taking logarithms we obtain

$$\theta(2n) - \theta(n) \leq 2n\log 2.$$

So we have that

$$\begin{aligned}\theta(2^k) - \theta(2^{k-1}) &\leq 2^k \log 2 \\ \theta(2^{k-1}) - \theta(2^{k-2}) &\leq 2^{k-1}\log 2 \\ &\vdots \\ \theta(2) - \theta(1) &< 2\log 2\end{aligned}$$

Summing up, $\theta(2^k) \leq 2^{k+1}\log 2$. Therefore for general n, it holds that

$$\theta(n) \leq \theta(2^{\lceil \log_2 n\rceil}) \leq 2^{\lceil \log_2 n\rceil + 1}\log 2 < 2^{\log_2 n + 2}\log 2 < 4n\log 2.$$

□

The following implies that $\psi(x) - \theta(x)$ is "small", I suggest you to try to prove this on your own.

Lemma 1.5 – $\psi(x) = \theta(x) + \mathcal{O}(\sqrt{x})$, and in particular $\psi(x) = \mathcal{O}(x)$.

Proof. We have that

$$\psi(x) = \theta(x) + \theta(x^{1/2}) + \theta(x^{1/3}) + \cdots$$

This summation is not infinite, we can write it as

$$\begin{aligned}\psi(x) = \theta(x) + \theta(x^{1/2}) + \sum_{k=3}^{\lceil \log_2 x \rceil} \theta(x^{1/k}) &= \theta(x) + \mathcal{O}(x^{1/2}) + \frac{\log x}{\log 2}\mathcal{O}(x^{1/3}) \\ &= \theta(x) + \mathcal{O}(\sqrt{x})\end{aligned}$$

This finally implies that $\psi(x) = \mathcal{O}(x)$ since $\theta(x) = \mathcal{O}(x)$. □

And we can now state the result we obtained initially as the proof is now complete:

Theorem 1.6 –

$$\sum_{n \leq x} \frac{\Lambda(n)}{n} = \log x + \mathcal{O}(1).$$

Recall the rough calculation we did for $\psi(x)$, it implied that $\psi(x)$ is roughly $\pi(x)\log x$. But we already got $\psi(x) = \mathcal{O}(x)$, so this should somehow imply estimates on $\pi(x)$ right? Yes, but with some work. Doing the calculations properly:

$$\begin{aligned}\psi(x) = \sum_{n \leq x} \Lambda(n) &= \sum_{p \leq x} \log p \left\lfloor \log_p x \right\rfloor \\ &= \sum_{p \leq x} \log p \left(\frac{\log x}{\log p} + \mathcal{O}(1) \right) \\ &= \pi(x)\log x + \mathcal{O}(1)\theta(x) \\ &= \pi(x)\log x + \mathcal{O}(x).\end{aligned}$$

Therefore we conclude that $\pi(x) = \mathcal{O}\left(\frac{x}{\log x}\right)$, this proves that there exist a positive constant b such that $\pi(x) < \frac{bx}{\log x}$, if you really want to you can get an explicit constant b easily, but this is not terribly important.

Remark 1.7. By elementary means you can show that $b = 1.6$ works.

We are left to prove a lower bound. There is a completely elementary proof[1] of the lower bound but here I will discuss a proof which uses the theory developed till now and also proves Mertens' first theorem on the fly.

Consider Theorem 1.6, for majority of the terms $n \leq x$, $\Lambda(n)$ is zero. And if n is not a prime but a power of a prime p, then the denominator of $\frac{\Lambda(n)}{n}$ becomes very large compared to the numerator, for those who know about p-series convergence it is probably immediate that the sum of terms when n is not a prime is bounded above by a constant, or in other words, $\mathcal{O}(1)$.

[1] https://math.stackexchange.com/a/1890792

Digression 1.8. Consider $S_p = \sum_{n=1}^{\infty} \frac{1}{n^p}$, called the $p-$series, it is well known that S_p converges for $p > 1$ and diverges for $p \leq 1$. Proving this is not hard, see here: https://math.stackexchange.com/a/29466

Theorem 1.9 Mertens' First Theorem **–**

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + \mathcal{O}(1)$$

Proof.

$$\begin{aligned}
\sum_{n \leq x} \frac{\Lambda(n)}{n} &= \sum_{p \leq x} \frac{\log p}{p} + \sum_{p^k \leq x, k \geq 2} \frac{\log p}{p^k} \\
&= \sum_{p \leq x} \frac{\log p}{p} + \sum_{p \leq x} \sum_{2 \leq k \leq \log_p x} \frac{\log p}{p^k} \\
&< \sum_{p \leq x} \frac{\log p}{p} + \sum_{p \leq x} \log p \sum_{k=2}^{\infty} \frac{1}{p^k} \\
&= \sum_{p \leq x} \frac{\log p}{p} + \sum_{p \leq x} \frac{\log p}{p^2 - p}
\end{aligned}$$

Note that $\sum_{n=2}^{\infty} \frac{\log n}{n^2 - n}$ is convergent because

$$\frac{\log n}{n^2 - n} < \frac{n^{0.1}}{n^2 - n} = \frac{1}{n^{1.9} - n^{0.9}} < \frac{1}{n^{1.8}}$$

holds for all sufficiently large n. We are now done by Theorem 1.6. □

Fix a large constant c. The above result implies that

$$\sum_{\frac{x}{c} < p \leq x} \frac{\log p}{p} = \sum_{p \leq x} \frac{\log p}{p} - \sum_{p \leq \frac{x}{c}} \frac{\log p}{p} = \log c + \mathcal{O}(1).$$

Here we pick c very large so that the RHS is positive, say the RHS is bounded below by $\delta > 0$. We do some trivial bounding,

$$(\pi(x) - \pi(x/c)) \frac{\log \frac{x}{c}}{x/c} \geq \sum_{\frac{x}{c} < p \leq x} \frac{\log p}{p} > \delta$$

Now just neglect $\pi(x/c)$ we get

$$\pi(x) > \delta \cdot \frac{x}{c(\log x - \log c)}$$

Thus we have established that there exists some a such that $\pi(x) > \frac{ax}{\log x}$. Again, with some hard work you can get an explicit constant but that is not very important.

Question 1.10. Put together a logical write-up of the proof of weak PNT.

Example 1.11 (Generalisation of Bertrand's Postulate)
Let $\varepsilon > 0$. Prove that there exist a prime between n and $(1+\varepsilon)n$ for all large n, in particular there always exist a prime between n and $2n$ for $n > 1$.

Demonstration. Just use PNT for the first part. Proving that there always exist a prime between n and $2n$ for $n > 1$ is doable[2] without the full power of PNT though. Hint: Consider $\binom{2n}{n}$.

Example 1.12
Fix $1 > \varepsilon > 0$. For some natural n, let $g(n)$ be the number of divisors of n in $(\sqrt{n}, (1+\varepsilon)\sqrt{n})$. Prove that $g : \mathbb{N} \to \mathbb{Z}_{\geq 0}$ is surjective.

Demonstration.

1. Check that $n = p^k$ for some prime p won't work.
2. Take $n = p^{2a}q^{2b}$ for two primes p and q.
3. Set some arbitrary k. You want to ensure that $p^a q^b < p^x q^y < (1+\epsilon)p^a q^b$ has k solutions in $0 \leq x \leq 2a, 0 \leq y \leq 2b$.
4. Fix x. At most how many possibilities for y are there?
5. It would be nice to have something like this: all such divisors are given by $p^a q^b (\frac{p}{q^t})^i$ for $i = 1, 2, \ldots, k$. Why do we expect this? When can this happen?
6. Obviously we want p/q^t to be very close to 1 and $tk = b$.
7. Finish using generalised Bertrand's postulate.

§1.3 Asymptotics for primes

We define p_n to be the nth prime number. It will be quite nice to to find a smooth function $f(n)$ such that $p_n \sim f(n)$. It turns out this is quite easy using PNT, the reader may try this on their own.

Theorem 1.13 — $p_n \sim n \log n$

Proof. Obviously $\pi(p_n) = n$. Therefore

$$\frac{p_n}{\log p_n} \sim n \implies \frac{p_n}{n \log n} \sim \frac{\log p_n}{\log n}.$$

But see that

$$n \sim \frac{p_n}{\log p_n} \implies \log n \sim \log p_n - \log\log p_n \implies \frac{\log p_n}{\log n} \sim 1 - \frac{\log\log p_n}{\log p_n}$$

Thus it follows that $p_n \sim n \log n$. □

This result is useful for ad-hoc calculations to get a feel about whether a statement or a conjecture should be true. Let me state a result without proof just to summarise:

[2] https://www.cut-the-knot.org/arithmetic/algebra/BertrandPostulate.shtml

Theorem 1.14 – The following are equivalent :

- $\pi(x) \sim x/\log x$
- $\theta(x) \sim x$
- $\psi(x) \sim x$
- $p_n \sim n \log n$

Remark 1.15 (Rosser's theorem)**.** $p_n > n \log n$

If you are interested then you may trying proving them. Finally here's a real olympiad problem:

Example 1.16 (EMMO 2016 Sr, Anant Mudgal)

We call a sequence of positive integers $\{a_n\}_{n\in\mathbb{N}}$ as a scouter if it is strictly increasing and $a_n < 9000n$. We call an integer $i \geq 1$ as divisor friendly if a_i divides the least common multiple of all previous terms of the sequence and call divisor-unfriendly otherwise. Is it necessarily true that a scouter has infinitely many

(a) divisor friendly

(b) divisor unfriendly

indices?

Demonstration.

1. Do part (b).
2. Assume that a_n is divisor unfriendly for all $n \geq N$.
3. Look at the prime factorisation of a_n, precisely, look at the exponents.
4. Conclude that there is a sequence of prime powers d_n such that $d_n \mid a_n$ and $d_i = d_j$ if and only if $i = j$.
5. Intuitively, d_n should grow faster than $9000n$.
6. Prove it by estimating the proportion of prime powers less than some large fixed number M. (We will discuss this idea in detail in the following section)
7. You may need to split the summation into the intervals $[1, \sqrt{9000n}]$ and $(\sqrt{9000n}, 9000n]$ and bound them separately, there are a lot of other ways to do this though.
8. Conclude.
9. Bonus: strengthen the bound. You can relax the upper bound for a_n to $\delta n \log n$ for some sufficiently small $\delta > 0$.

§2 Density

§2.1 Asymptotic Density

Density of a subset S of $\mathbb{N}$ refers to the "proportion" of positive integers which are in S. For example, what is the density of even numbers? Or in other words, what proportion of positive integers are even numbers? Intuitively, the answer is $\frac{1}{2}$. This notion is captured formally as:

Definition 2.1. Let S be a set of positive integers. The **asymptotic density** of S is defined as

$$d(S) = \lim_{n\to\infty} \frac{|S \cap \{1, 2, \ldots, n\}|}{n}$$

if the limit exists.

This may feel like a mouthful, but what the definition says is that we find the proportion for a finite n and then take the limit $n \to \infty$. One can also think of this as the probability that a positive integer chosen at random belongs to S. What is the density of the set of prime numbers? Zero.

The following example showcases the power of density :

Example 2.2 (China TST 3 2015/3)

Prove that there exist infinitely many integers n such that $n^2 + 1$ is square-free.

Demonstration. The main idea is to estimate the number of positive integers $n \le N$ such that $n^2 + 1$ is square-free using truncated Inclusion Exclusion Principle for some fixed N.

1. Define $A_N = \{n^2+1 \mid n \le N,\ n^2+1 \text{ is square-free}\}$. Which primes divide numbers of the form $n^2 + 1$?

2. Fix some odd prime $p \equiv 1 \pmod 4$. At most how many multiples of p^2 are in A_N? To find this, first show that the congruence $n^2 + 1 \equiv 0 \pmod{p^2}$ has at most 2 solutions modulo p^2.

3. What happens when $p = 2$? How many multiples of 4 are there in A_N?

4. Show that number of non-squarefree numbers in A_N is at most

$$\sum_{p\le N,\ 4|p-1} 2\left\lceil \frac{N}{p^2} \right\rceil \le 2\sum_{p\le N}\left(\frac{N}{p^2}+1\right) = 2N\sum_{p\le N}\frac{1}{p^2} + \mathcal{O}\left(\frac{N}{\log N}\right)$$

   Notice that we are totally dropping $p \equiv 1 \pmod 4$ for now.

5. Show that the proportion (density) of square-free numbers in A_N is at least

$$1 - 2\sum_{p\le N}\frac{1}{p^2} - \mathcal{O}\left(\frac{1}{\log N}\right) \sim 1 - 2\sum_{p\le N}\frac{1}{p^2}.$$

6. Conclude by proving that

$$\sum_{p}\frac{1}{p^2} < \frac{1}{2}.$$

Remark 2.3. By tightly bounding, you can show that the density of n such that n^2+1 is square-free is at least 0.8924. This means that a positive integer of the form n^2+1 picked at random has at least 89.24% chances of being square-free!

Remark 2.4. Note that here we are not proving the existence of the limit for density, it will be painful to prove the existence every time we want to talk about density. Often we only care about bounds on the density rather than computing its exact value. To take care of this issue we define

$$d_{\text{upper}}(S) = \lim_{n\to\infty} \sup \frac{|S \cap \{1,2,\ldots,n\}|}{n}$$

called the **Upper Density** of the set S and

$$d_{\text{lower}}(S) = \lim_{n\to\infty} \inf \frac{|S \cap \{1,2,\ldots,n\}|}{n}$$

called the **Lower Density** of the set S. So density of a set S exists if and only if $d_{\text{upper}}(S) = d_{\text{lower}}(S)$. If you want to be fully rigorous you can replace every word "density" with whatever seems suitable from the above two in the rest of this article.

Digression 2.5. Define

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s},$$

called the **Riemann Zeta Function**. It is well known that $\zeta(2) = \frac{\pi^2}{6}$ (buzzword: "Basel problem"), the reason I am introducing this is because a few contest problems require you to bound the sum of reciprocals of squares of primes, or in general sum of reciprocals of squares of some set of positive integers. You can also prove that

$$\zeta(s) = \prod_{p} \left(1 - \frac{1}{p^s}\right)^{-1},$$

this is known as the **Euler Product Formula**.

§2.2 Kronecker's Theorem and Equidistribution Theorem

This section is a bit dense (no pun intended) so take your time.

You may already know what dense means, if not, here is the definition (note that denseness is defined far more generally, here we only consider $\mathbb{R}$) :

Definition 2.6. Let A be a subset of $S \subseteq \mathbb{R}$, we say that A is **dense** in S if for every $x \in S$ and $\varepsilon > 0$, there exists an element $a \in A$ such that $a \in (x-\varepsilon, x+\varepsilon)$.

For our purposes S will mostly be an interval. Suppose $S = [0,1]$ and let $A \subseteq S$ be dense in S. One can think of this as - there are elements of A arbitrarily close to both 0 and 1 and for any $a, b \in A$, there is some $c \in (a,b)$ which belongs to A. For example, the set of rational numbers in $[0,1]$ is dense in $[0,1]$. The way I like to think about this (in case of intervals of $\mathbb{R}$ of course) is that between any two elements of A there exists another element of A.

Theorem 2.7 Kronecker's Theorem **—** Let k be an irrational number. The set of fractional parts of the terms of the sequence $\{nk\}_{n=1}^{\infty}$ is dense in $[0,1]$.

Proof. This is not difficult. Left as an exercise. See the end of the article for a proof. □

Actually, far more is true about the sequence $\{nk \pmod 1\}_{n=1}^{\infty}$ (here we write $\pmod 1$ to denote fractional parts, quite self-explanatory) where k is irrational, it isn't only dense in $[0,1]$ but "uniformly" dense in $[0,1]$. "Uniformly" dense is exactly what you think it means - distributed evenly. So we can say that equidistribution is nicer than simply being dense. This is defined formally as :

Definition 2.8. Let $\{a_n\}_{n\geq 1}$ be a sequence of real numbers in the interval $[0,1]$. We say that the sequence is **equidistributed** if

$$\lim_{n\to\infty} \frac{|\{i \mid 1 \leq i \leq n,\ a_i \in [a,b]\}|}{n} = b - a$$

holds for all real numbers $0 \leq a \leq b \leq 1$.

Question 2.9. Digest the above definition. Prove that Equidistributed $\implies$ Dense.

Theorem 2.10 Equidistribution Theorem **–** Let k be an irrational number. The sequence of fractional parts of the terms of the sequence $\{nk\}_{n=1}^{\infty}$ is equidistributed in $[0,1]$.

Demonstration.

1. By Kronecker's theorem there exists $N \in \mathbb{N}$ such that $\{Nk\} < \varepsilon$ for some very small $\varepsilon > 0$.
2. Consider the sequence T :
$$\{Nk\}, \{2Nk\}, \{3Nk\}, \ldots$$
3. Imagine a number line and consider the interval $[0,1]$. Plot the sequence T term by term on the number line. Observe that there will be continuous runs of terms which belong to $I = [a,b]$ separated by runs of terms which don't belong to I.
4. About how long are the runs of terms of both types?
5. Fix some large M. What proportion of the first M terms are in I?
6. Do the same thing with
$$\{ik\}, \{(i+N)k\}, \{(i+2N)k\}, \ldots$$
for all $1 \leq i < N-1$.
7. Sum up and conclude.

See the end of the article for a complete proof.

Example 2.11

Find the (asymptotic) density of positive integers n such that 7^n begins with the digits 42 in base-10.

Demonstration.

1. Prove that this is equivalent to having

$$\log_{10} \frac{43}{10} > \{n \log_{10} 7\} \geq \log_{10} \frac{42}{10}$$

2. Using Equidistribution theorem, $\{n \log_{10} 7\}$ is equidistributed in $[0, 1]$. Using the definition of equidistribution, conclude that the required density is

$$\log_{10} \frac{43}{10} - \log_{10} \frac{42}{10} = \log_{10} \frac{43}{42}.$$

§3 Mertens' Second Theorem

Few readers may already that the sum of reciprocals of primes is divergent. It is true that

Theorem 3.1 weak form of Mertens' Second Theorem **—**

$$\sum_{p \leq n} \frac{1}{p} = \log \log n + \mathcal{O}(1)$$

Question 3.2. For those who know integration, why do we expect the LHS to be asymtotic to $\log \log n$?

See the last page for the proof of one direction.

§4 PNT for Arithmetic Progressions

This section unfortunately will be lack proofs because they are not in the scope of olympiad mathematics, but these results are very nice, so I decided to include them. We state a marvellous result without proof :

Theorem 4.1 PNT for Arithmetic Progressions **—** Let r and d be two relatively prime positive integers. The number of primes less than x which are congruent to r modulo d is asymptotic to

$$\frac{1}{\varphi(d)} \cdot \frac{x}{\log x}$$

.

One can kind of intuitively see why this should be true - there are $\varphi(d)$ invertible residues modulo d, namely, those coprime to d. Almost all the other theorems and estimates change the way you would expect, they are scaled down by $\varphi(d)$.

Theorem 4.2 — If a and d are relatively prime positive integers then

1.

$$\sum_{\substack{p \leq x \\ p \equiv a \pmod{d}}} \frac{\log p}{p} \sim \frac{1}{\varphi(d)} \cdot \log x$$

2. $$\sum_{\substack{p^k \le x \text{ for some } k\in\mathbb{N} \\ p\equiv a \pmod d}} \log p \sim \frac{1}{\varphi(d)} \cdot x$$

3. $$\sum_{\substack{p\le x \\ p\equiv a \pmod d}} \log p \sim \frac{1}{\varphi(d)} \cdot x$$

4. $$\sum_{\substack{p\le x \\ p\equiv a \pmod d}} \frac{1}{p} \sim \frac{1}{\varphi(d)} \cdot \log\log x$$

Question 4.3. Why do we expect the scaling by $\frac{1}{\varphi(d)}$?

For further references on this, see [4] or [5].

§5 Examples

I feel that this topic requires more examples than usual so plenty of examples follow from here, if you get bored feel free to skip to the latter examples or next sections :)

Example 5.1
Prove that the sequence $\{\lfloor p_n\nu \rfloor\}_{n=1}^{\infty}$ has infinitely many prime divisors where ν is some positive real number greater than 1.

Demonstration. The main idea here is to look at sum of reciprocals.

1. Show that if $\{a_n\}_{i=1}^{\infty}$ is a sequence which has finitely many prime divisors say q_i for $i = 1, 2, \ldots, k$, then $\sum \frac{1}{a_n}$ is convergent. Use the crude bound:
$$\sum_{i=1}^{\infty} \frac{1}{a_n} \le \prod_{i=1}^{k} \left(1 + \frac{1}{q_i} + \frac{1}{q_i^2} + \cdots\right).$$
2. Conclude.

Example 5.2 (Mathlinks)
Find all polynomials $p(x) \in \mathbb{Z}[x]$ such that for all positive integers n, we have that $p(n)$ is a palindrome number.

A number q written in base 10 is called a Palindrome number, if q reads the same from left to right, as it reads from right to left. For example : $121, -123321$ are Palindrome numbers, but 113 is not a Palindrome number.

Demonstration.

1. Suppose p is non-constant. Let $d = \deg p > 1$.

2. It is pretty intuitive that there exist arbitrarily long runs of consecutive natural numbers whose $p-$values start with the same fixed digit.
3. Verify it using $p(x) \sim ax^d$ where a is the coefficient of x^d.
4. Finish with modulo 10.

Example 5.3 (Canada MO 2020/4)
Let $S = \{1, 4, 8, 9, 16, \ldots\}$ be the set of all perfect powers i.e. $S = \{n^k \mid n, k \in \mathbb{Z}, k \geq 2\}$. We arrange the elements of S into an increasing sequence $\{a_i\}_{i=1}^{\infty}$. Show that there are infinite many positive integers n such that $9999 \mid a_{n+1} - a_n$.

Demonstration.

1. Find pairs of consecutive perfect squares whose difference is divisible by 9999. Parametrize to get many such pairs.
2. You want to show that there are infinitely many such pairs of consecutive perfect squares between which there is no perfect power.
3. Verify that these "pairs" of perfect squares are far denser than perfect odd powers to conclude.

Example 5.4 (Putnam 2007)
Find all polynomials f with real coefficients such that if n is a positive integer which is written in base 10 only with ones, then $f(n)$ has the same property.

Demonstration.

1. Let $f(\frac{10^n-1}{9}) = \frac{10^{x_n}-1}{9}$ for all n where x_n is a sequence of positive integers.
2. Suppose f is non-constant. Let the degree of f be $d \geq 1$ and $f(x) = ax^d + o(x^d)$.
3. So it follows that
$$f\left(\frac{10^n - 1}{9}\right) \sim \frac{a \cdot 10^{nd}}{9^d}.$$
4. Show that the sequence $x_n - nd$ is convergent. Let the limit be L then see that $a = 9^{d-1} \cdot 10^L$.
5. Observe that $x_n - nd$ must be eventually constant.
6. Finish the problem.

Example 5.5 (USAMO 2014/6)
Prove that there is a constant $c > 0$ with the following property: If a, b, n are positive integers such that $\gcd(a + i, b + j) > 1$ for all $i, j \in \{0, 1, \ldots, n\}$, then
$$\min\{a, b\} > c^n \cdot n^{\frac{n}{2}}.$$

Demonstration.

1. Make an $(n+1) \times (n+1)$ table with the i,jth entry being the smallest prime divisor of $\gcd(a+i, b+j)$.
2. Try filling up the table with primes. Observe that the primes get large really quick.
3. Take some prime p. Get an upper bound on the number of times p can appear in the table.
4. Fix some large C. Show that the maximum number of entries that are occupied by primes at most C is something like
$$\sum_{p \le C} \left\lceil \frac{n+1}{p} \right\rceil^2.$$
5. Do some bounding and conclude that there exist a constant $c > 0$ such that at least 50% of the primes in the table are larger than cn^2.
6. Thus there is some row/column with at least half of its primes larger than cn^2.
7. Conclude.

Example 5.6 (Iran 3rd round 2011)
Suppose that α is a real number and $a_1 < a_2 < \cdots$ is a strictly increasing sequence of natural numbers such that for each natural number n we have $a_n \le n^\alpha$. We call the prime number q golden if there exists a natural number m such that $q | a_m$. Suppose that $q_1 < q_2 < q_3 < \cdots$ are all the golden prime numbers of the sequence $\{a_n\}$.

(a) Prove that if $\alpha = 1.5$, then $q_n \le 1390^n$. Can you find a better bound for q_n?

(b) Prove that if $\alpha = 2.4$, then $q_n \le 1390^{2n}$. Can you find a better bound for q_n?

Demonstration. This problem is quite tricky. We only demonstrate part (a), part (b) is similar.

1. Assume the contrary that there exist $q_n > 1390^n$, and take n to be minimal. Suppose r is the minimal index such that $q_n \mid a_r$.
2. Get a lower bound on
$$S = \sum_{k=1}^{r-1} \frac{1}{a_k^{\frac{1}{3}}}$$
just by using $a_n \le n^{1.5}$.
3. Using the fact that all prime factors of the elements of the set $\{a_1, a_2, \ldots, a_{r-1}\}$ belong to the set $\{q_1, q_2, \ldots, q_{n-1}\}$, get an upper bound on S.
4. Combine and conclude.
5. Strengthen the bound.
6. Try considering $\sum_{i=1}^{r-1} \frac{1}{a_k}$ or $\sum_{i=1}^{r-1} \frac{1}{a_k^{0.5}}$, what happens? You may also try taking $\sum_{i=1}^{r-1} a_k^{-s}$ for some unspecified s.

Example 5.7 (IMO 2008/3 improved)
Let $\varepsilon > 0$. Prove that there exist infinitely many n such that there is a prime divisor of $n^2 + 1$ which is larger than $(1 - \varepsilon)n \log n$.

Demonstration. This is a pretty hard problem if you are not familiar with some common methods and ideas in this subject.

1. The key idea is to analyse the product $f(N) = \prod_{n=1}^{N}(n^2 + 1)$ (note that this is the exact same "global" idea discussed in the beginning of the Section 1.2).
2. If $p \leq N$ is a $1 \pmod 4$ prime then show that
$$\nu_p(f(N)) \leq 2\left\lceil \frac{N}{p} \right\rceil + 2\left\lceil \frac{N}{p^2} \right\rceil + 2\left\lceil \frac{N}{p^3} \right\rceil + \cdots + 2\left\lceil \frac{N}{p^k} \right\rceil$$
where $k = \lceil \log_p N \rceil$. Handle $p > N$ and $p = 2$ separately.
3. Get an upper bound on $f(N)$ which looks something like
$$\log f(N) \leq 2 \log N \sum_{\substack{p \leq N \\ p \equiv 1 \pmod 4}} 1 + 2N \sum_{\substack{p \leq N \\ p \equiv 1 \pmod 4}} \frac{\log p}{p-1} + \sum_{\substack{p \leq t \\ p \equiv 1 \pmod 4}} \log p + \frac{N}{2} \log 2$$
where t is the largest prime divisor of $f(N)$.
4. Bound $f(N)$ from below and conclude.

Remark 5.8. Let $c > 0$ be a sufficiently small constant. You can try showing that the set of all positive integers n such that $n^2 + 1$ has a prime divisor larger than $cn \log n$ has positive density.

Remark 5.9. In fact, there exist infinitely many n such that $n^2 + 1$ has a prime divisor larger than $n^{6/5}$. The proof is non-elementary.

Example 5.10 (STEMS 2020 B3/C5, Arka Karmakar)
Let $S(a) := \{a^i + a^j \mid i, j \in \mathbb{N}\}$. Find all tuples (a, b, f) with $a, b \in \mathbb{N}$ and $f \in \mathbb{R}[x]$ such that $f(S(a)) \subseteq S(b)$. (Here $f(S(a))$ denotes the image of $S(a)$ under f.)

Demonstration. This one is a pretty hard and the solution is a bit involved as well.

1. Guess the solutions. Note that the leading coefficient of f must be positive and $f \in \mathbb{Q}[x]$ (Prove it). For now assume that f is non-constant.
2. Consider $f(a^n + a) = b^{x_1} + b^{y_1}$ and $f(a^n + a^2) = b^{x_2} + b^{y_2}$ where $x_1 \geq y_1$ and $x_2 \geq y_2$. We expect $x_1 = x_2$ for all large n. Why?
3. Prove the above. Hint: For $n \to \infty$ the ratio of $f(a^n + a)$ and $f(a^n + a^2)$ converges to 1, this is basically the main intuition rigorised. Make separate cases for $b > 2$ and $b = 2$ if needed.
4. Generalise/extend the above.

5. Fix k, pick huge n and let $f(a^n + a^i) = b^{t_n} + b^{A(n,i)}$ for all $i = 1, 2, 3, \ldots, k$. Here we can let it to be t_n since it is independent of i for small i.

6. It is not a bad guess that $A(n, i)$ is an arithmetic progression for small i, you should actually expect this to be true if you have already guessed the solutions.

7. To prove this, analyse ν_p for some prime p. Writing $f(x) = x^d(xg(x) + c)$ might be helpful (you secretly know what d and c should be).

8. You would probably need this as a lemma: If $p \mid b$ then $p \mid a$ for all primes p (Prove it)

9. Finish the problem.

§6 Problems

All the problems below don't necessarily use the theory discussed in this article. Many of the following problems are hard so don't get demotivated.

§6.1 Exercises

If you are experienced then you may skip this section.

Exercise 6.1. Let A be a set of positive integers with positive asymptotic density. Prove that sum of reciprocals of elements of A is divergent.

Exercise 6.2. If the density of $A \subset \mathbb{N}$ and $B \subset \mathbb{N}$ is zero then prove that density of $A \cup B$ is zero.

Exercise 6.3. You are given a string of base-10 digits. Prove that you can append some finite number of digits so that the resultant number becomes a power of 2.

Exercise 6.4. Define $\omega(n)$ to be the number of distinct prime divisors of n. Prove that

$$\sum_{n \le x} \omega(n) = x \log \log x + \mathcal{O}(x).$$

Exercise 6.5. Prove that

$$\prod_p \left(1 - \frac{1}{p}\right) = 0$$

where the product is over all primes p.

Exercise 6.6. Let $r_2(n)$ be the number of ways n can be written as a sum of two perfect squares. Prove that

$$\lim_{n \to \infty} \frac{r_2(1) + r_2(2) + \cdots + r_2(n)}{n} = \pi.$$

Exercise 6.7 (Mathotsav)**.** We say that a positive integer t is good if the density of positive integers n such that $n^2 + t$ is square-free is at least 0.99.

(a) Prove that the density of square free numbers is $\frac{6}{\pi^2}$.

(b) Prove that infinitely many natural numbers are good.

(c) Prove that there exists a positive constant c and a natural number N, such that for all $n > N$, the number of natural numbers less than n which are good is at least cn.

§6.2 Easy

Problem 6.8 (Iranian Our MO 2020)**.** Consider two sequences $x_n = an + b$, $y_n = cn + d$ where a, b, c, d are natural numbers and $\gcd(a, b) = \gcd(c, d) = 1$, prove that there exist infinite n such that x_n, y_n are both square-free.

Problem 6.9 (Iran 3rd round 2010/8)**.** Prove that there are infinitely many natural numbers of the form $n^2 + 1$ such that they don't have any divisor of the form $k^2 + 1$ except 1 and themselves.

Problem 6.10 (China TST 2005)**.** Prove that for any n $(n \geq 2)$ pairwise distinct fractions in the interval $(0, 1)$, the sum of their denominators is no less than $\frac{1}{3}n^{\frac{3}{2}}$.

Problem 6.11 (China TST 2004)**.** Let u be a fixed positive integer. Prove that the equation $n! = u^\alpha - u^\beta$ has a finite number of solutions (n, α, β).

Problem 6.12 (IMO Shortlist 2011/A2)**.** Determine all sequences $(x_1, x_2, \ldots, x_{2011})$ of positive integers, such that for every positive integer n there exists an integer a with

$$\sum_{j=1}^{2011} j x_j^n = a^{n+1} + 1.$$

Problem 6.13 (China TST 2010, Miklos Schweitzer, Paul Erdos)**.** Given positive integers n and k such that $n \geq 9^k$, prove that $\binom{n}{k}$ has at least k different prime divisors.

Problem 6.14 (IMO ShortList 2003/N4)**.** Let b be an integer greater than 5. For each positive integer n, consider the number

$$x_n = \underbrace{11\cdots1}_{n-1}\underbrace{22\cdots2}_{n}5,$$

written in base b.

Prove that the following condition holds if and only if $b = 10$: there exists a positive integer M such that for any integer n greater than M, the number x_n is a perfect square.

Problem 6.15 (Vesselin Dmitrov)**.** Prove that the set of positive integers n such that

$$\tfrac{1}{2}n(n+1)(n+2)(n^2+1)$$

is square free has positive density.

Problem 6.16 (Miklos Schweitzer)**.** Prove that the set of positive integers n such that $\tau(n) \mid n$ has density 0.

§6.3 Medium

Problem 6.17 (ARMO 2012 Grade 11 Day 2)**.** For a positive integer n define $S_n = 1! + 2! + \ldots + n!$. Prove that there exists an integer n such that S_n has a prime divisor greater than 10^{2012}.

Problem 6.18 (AoPS)**.** Prove that $n! = m^3 + 8$ has only finitely many solutions in positive integers.

Problem 6.19 (China TST 2 Day 1 P1)**.** Let n be a positive integer. Let D_n be the set of all divisors of n and let $f(n)$ denote the smallest natural m such that the elements of D_n are pairwise distinct in mod m. Show that there exists a natural N such that for all $n \geq N$, one has $f(n) \leq n^{0.01}$.

Problem 6.20. (Own & `Superguy`, KöMaL A. 787) Let p_n denote the n^{th} prime number and define $a_n = \lfloor p_n \nu \rfloor$ for all positive integers n where ν is a positive irrational number. Is it possible that there exist only finitely many k such that $\binom{2a_k}{a_k}$ is divisible by p_i^{10} for all $i = 1, 2, \ldots, 2020$?

Problem 6.21 (`Superguy`)**.** Prove that the set of positive integers n such that n and $2^n - 1$ are relatively prime has positive lower density.

Problem 6.22 (Paul Erdos, Miklos Schweitzer)**.** Let $a_1 < a_2 < \cdots < a_n$ be a sequence of positive integers such that $a_i - a_j \mid a_i$ for all $i \leq j$. Prove that there is a positive constant c such that for any such sequence of length n, $a_1 > n^{cn}$.

Problem 6.23. (Tuymaada 2011, Senior Level) Let $P(n)$ be a quadratic trinomial with integer coefficients. For each positive integer n, the number $P(n)$ has a proper divisor d_n, i.e., $1 < d_n < P(n)$, such that the sequence $d_1, d_2, d_3, \ldots$ is increasing. Prove that either $P(n)$ is the product of two linear polynomials with integer coefficients or all the values of $P(n)$, for positive integers n, are divisible by the same integer $m > 1$.

Problem 6.24. (Turkey TST 2015/6) Prove that there are infinitely many positive integers n such that $(n!)^{n+2015}$ divides $(n^2)!$.

Problem 6.25 (China TST 2015)**.** Let $a_1, a_2, a_3, \ldots$ be distinct positive integers, and $0 < c < \frac{3}{2}$. Prove that : There exist infinitely many positive integers k, such that $\text{lcm}(a_k, a_{k+1}) > ck$.

> **Remark 6.26.** The bound cannot be improved to $\text{lcm}(a_k, a_{k+1}) > k^{1+\delta}$ for some $\delta > 0$.

Problem 6.27 (USA TSTST 2017/6)**.** A sequence of positive integers $(a_n)_{n \geq 1}$ is of Fibonacci type if it satisfies the recursive relation $a_{n+2} = a_{n+1} + a_n$ for all $n \geq 1$. Is it possible to partition the set of positive integers into an infinite number of Fibonacci type sequences?

Problem 6.28 (Tuymaada 2007/8)**.** Prove that there exists a positive c such that for every positive integer N among any N positive integers not exceeding $2N$ there are two numbers whose greatest common divisor is greater than cN. (Bonus: Strengthen the bound)

§6.4 Hard

Problem 6.29 (IMO 2015/N6)**.** Let $\mathbb{Z}_{>0}$ denote the set of positive integers. Consider a function $f : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$. For any $m, n \in \mathbb{Z}_{>0}$ we write $f^n(m) = \underbrace{f(f(\ldots f}_{n}(m)\ldots))$.

Suppose that f has the following two properties:

(i) if $m, n \in \mathbb{Z}_{>0}$, then $\frac{f^n(m) - m}{n} \in \mathbb{Z}_{>0}$;

(ii) The set $\mathbb{Z}_{>0} \setminus \{f(n) \mid n \in \mathbb{Z}_{>0}\}$ is finite.

Prove that the sequence $f(1) - 1, f(2) - 2, f(3) - 3, \ldots$ is periodic.

Problem 6.30 (China TST 2018 Day 2 Q2)**.** Given a positive integer k, call n good if among

$$\binom{n}{0}, \binom{n}{1}, \binom{n}{2}, \ldots, \binom{n}{n}$$

at least $0.99n$ of them are divisible by k. Show that exists some positive integer N such that among $1, 2, \ldots, N$, there are at least $0.99N$ good numbers.

Problem 6.31 (Paul Erdos)**.** For any $\delta > 0$ prove that there are at least $(\frac{2}{3} - \delta)\frac{n}{\log_2 n}$ primes between n and $2n$ for sufficiently large n. (Using the full power of PNT would be cheating :P)

Problem 6.32 (IMO Shortlist 2019/N7)**.** Prove that there is a constant $c > 0$ and infinitely many positive integers n with the following property: there are infinitely many positive integers that cannot be expressed as the sum of fewer than $cn\log(n)$ pairwise coprime nth powers.

Problem 6.33 (XIII Brazilian Olympic Revenge 2014)**.** Let $a > 1$ be a positive integer and $f \in \mathbb{Z}[x]$ with positive leading coefficient. Let S be the set of integers n such that

$$n \mid a^{f(n)} - 1.$$

Prove that S has density 0; that is, prove that $\lim_{n\to\infty} \frac{|S\cap\{1,\ldots,n\}|}{n} = 0$.

Problem 6.34 (PRIMES 2020 M5)**.** We say an integer $n \geq 2$ is chaotic if for any monic nonconstant polynomial $f(x)$ with positive integer coefficients, the set

$$\{f(1), f(2), \ldots, f(n)\}$$

contains fewer than $10^{\deg f} \cdot \frac{n}{\log n}$ prime numbers. Are there finitely many chaotic integers?

> **Remark 6.35.** There is a theorem by Nagell & Heilbronn which says that for any $f \in \mathbb{Z}[x]$, the number of primes in $\{|f(1)|, |f(2)|, \ldots, |f(n)|\}$ is $\mathcal{O}(n/\log n)$ but unfortunately the proof is beyond the scope of olympiad Mathematics.

Problem 6.36 (Marius Cavachi, AMM)**.** Let a and b be integers greater than 1 such that $a^n - 1 \mid b^n - 1$ for every positive integer n. Prove that b is a natural power of a.

> **Remark 6.37.** You can relax the condition to "for infinitely many positive integers n" instead of "for every positive integer n" and the problem would still hold. However the proof is non-elementary.

Problem 6.38 (Fedor Petrov)**.** Does there exist $c > 0$ such that among any n positive integers one may find 3 with least common multiple at least cn^3?

§7 Solutions to selected examples

§7.1 Example 1.12

Let's suppose we want $g(n) = k$. Choose a large enough natural t such that $2^t < q < 2^t(1+\varepsilon)$ where q is a prime. Note that $n = q^{2k}2^{2kt}$ works because all such k divisors are of the form $q^{k+i}2^{t(k-i)}$ for $i = 1, 2, \ldots, k$. No other divisor works because for any fixed power of q we can have only one power of 2 which may work.

§7.2 Example 1.15 (EMMO 2016 Sr, Anant Mudgal)

Part (b) is easy so we only solve part (a). Assume the contrary that all sufficiently large indices are divisor friendly. We have that

$$a_n \nmid \operatorname{lcm}(a_1, a_2, \ldots, a_{n-1})$$

for all $n > K$, say. Observe that there must exist some sequence of primes q_n such that if $b_n = q_n^{\nu_{q_n}(a_n)}$ for $n > K$ then q_n divides none of the preceding terms a_i for $i > K$. See that all the b_n's must be distinct. Obviously $b_1, b_2, \ldots, b_n \le 9000n$ and b_i are distinct prime powers. Number of prime powers at most $9000n$ is less than

$$\begin{aligned}
S = \sum_{p \le 9000n} \log_p 9000n &= \log 9000n \sum_{p \le 9000n} \frac{1}{\log p} \\
&\le \log 9000n \left(\sum_{p < \sqrt{9000n}} \frac{1}{\log p} + \sum_{\sqrt{9000n} \le p \le 9000n} \frac{1}{\log p} \right) \\
&\le \log 9000n \left(\frac{\sqrt{9000n}}{\log 2} + \frac{1}{\log \sqrt{9000n}} \cdot (\pi(9000n) - \pi(\sqrt{9000n})) \right) \\
&= \log 9000n \left(\frac{\sqrt{9000n}}{\log 2} + \frac{1}{\log \sqrt{9000n}} \cdot \mathcal{O}\left(\frac{n}{\log n} \right) \right) \\
&= \log 9000n \cdot \mathcal{O}\left(\frac{n}{\log^2 n} \right) = \mathcal{O}\left(\frac{n}{\log n} \right),
\end{aligned}$$

in the last second step we used PNT. This is a contradiction for large enough n since there are n distinct prime powers at most $9000n$, namely $b_1, b_2, \ldots, b_n$. And we are done.

§7.3 Example 5.3 (Canada MO 2020/4)

Consider $(9999n + 4999)^2$ and $(9999n + 5000)^2$, verify that their difference is divisible by 9999, call a pair of such perfect squares good. Fix some large N. Check that the number of such pairs less than N is bounded below by $c\sqrt{N}$ for some constant $c > 0$. All perfect powers between such a pair must be odd perfect powers. Number of odd perfect powers a^b less than N is at most

$$S = N^{1/3} + N^{1/5} + N^{1/7} + \cdots$$

where the number of summands is at most $\log_2 N$ as $a \ge 2$ except for the trivial perfect power 1. Therefore $S = \mathcal{O}(N^{1/3} \log N)$, which is less than $c\sqrt{N}$ for all large N. Thus there exists infinitely many good pairs.

§7.4 Example 5.6 (Iran 3rd round 2011)

(Solution by `Superguy`) We are going to prove the bound $q_n \leq 35^n$ for part (a). Let's assume for the sake of contradiction that there exists n such that $q_n > 35^n$, here suppose n is minimal.Then suppose r is the minimal index such that $q_n \mid a_r$ then $r > 35^{\frac{2}{3}n}(\clubsuit)$. So all of $\{a_1, a_2, \ldots, a_{r-1}\}$ have prime factors in set $\{q_1, q_2, \ldots, q_{n-1}\}$. Call this set of primes as P. We clearly have

$$\sum_{k=1}^{r-1} \frac{1}{a_k^{\frac{1}{3}}} \geq \sum_{k=1}^{r-1} \frac{1}{\sqrt{k}} \qquad (1)$$

Clearly RHS in (1) is greater than $2\sqrt{r} - 2$ which can be shown using easy integration or induction. Consider the following claim:

Claim –

$$\sum_{k=1}^{r-1} \frac{1}{a_k^{\frac{1}{3}}} \leq \prod_{p \in P} \left[\sum_{m \geq 0} p^{-\frac{1}{3}m} \right] \leq 5(3.27)^{n-2}$$

where $|P| = n - 1$.

Proof. Note that all of a_k are of the form $q_1^{k_1} \cdot q_2^{k_2} \cdots q_{n-1}^{k_{n-1}}$ where all k_i are non-negative which gives the left side inequality. For right side we have that the sum $\sum_{m \geq 0} p^{-\frac{1}{3}m}$ is maximum for $p = 2$ and next greatest value is achieved by $p = 3$ and the value of the sum for $p = 2$ is less than 5 and for $p = 3$ the sum would be less than 3.27 Now observe

$$\prod_{p \in P} \left[\sum_{m \geq 0} p^{-\frac{1}{3}m} \right] \leq 5(3.27)^{n-2}$$

So we get the claim. □

Now by our claim,(1),($\clubsuit$) and one fact:

$$\ln(2 \cdot 35^{\frac{n}{3}}) < \ln(2 \cdot 35^{\frac{n}{3}} - 2) + 1 \text{ for all natural } n,$$

we get that we should have

$$\ln(2) - 1 + \frac{n \cdot \ln(35)}{3} - \ln(5) - (n-2)(\ln(3.27)) < 0 \qquad (2)$$

Now we are going to prove the opposite inequality. Taking the function in LHS as $f(n)$ we get that $f(n)$ is increasing. Hence we just need to check for $n = 1$ which we get that $f(1) > 0$. Thus we have proved the opposite inequality and thus the contradiction for our initial assumption. For part b exact similar process can give a nice bound of some $q_n < 300^n$. ■

§7.5 Example 5.7 (IMO 2008/3 improved)

Define

$$f(N) = \prod_{n \leq N} (n^2 + 1).$$

Lets assume that the largest prime divisor of $f(N)$ is t. Let $f(N) = \prod_p p^{\alpha_p}$ be the prime factorisation of $f(N)$, each prime $p > N$ can divide $n^2 + 1$ for at most two different values of n, and so $\alpha_p \leq 2$ in this case. See that $\alpha_2 = \lfloor N/2 \rfloor$. For $p \leq x$, if $p \mid n^2 + 1$,

then $n^2 \equiv -1 \bmod p$ which has solutions if and only if $p \equiv 1 \pmod 4$, and in that case there will be at most $2\lceil N/p \rceil$ values of n for which $p \mid n^2+1$. Similarly, if $p^k \mid n^2+1$, then $n^2 \equiv -1 \bmod p^k$, and there are at most 2 solutions to this congruence and hence at most $2\lceil N/p^k \rceil$ values of n for which $p^k \mid n^2+1$. Combining, we find that for $p \le N$ and $p \equiv 1 \pmod 4$

$$\alpha_p \le 2\left\lceil \frac{N}{p} \right\rceil + 2\left\lceil \frac{N}{p^2} \right\rceil + 2\left\lceil \frac{N}{p^3} \right\rceil + \cdots + 2\left\lceil \frac{N}{p^k} \right\rceil$$

where $k = \lceil \log_p N \rceil$. This gives that

$$\alpha_p \le \frac{2N}{p-1} + 2\left(\log_p N + 1\right)$$

since $1 + \frac{1}{p} + \frac{1}{p^2} + \cdots + \frac{1}{p^k} \le \frac{1}{1-1/p}$. Thus,

$$f(N) \le 2^{N/2} \prod_{\substack{p \le N \\ p \equiv 1 \pmod 4}} p^{\frac{2N}{p-1} + 2\log_p(N) + 2} \prod_{\substack{N < p \le t \\ p \equiv 1 \pmod 4}} p^2,$$

and so,

$$f(N) \le 2^{N/2} \prod_{\substack{p \le N \\ p \equiv 1 \pmod 4}} N^2 \prod_{\substack{p \le N \\ p \equiv 1 \pmod 4}} p^{\frac{2N}{p-1}} \prod_{\substack{p \le t \\ p \equiv 1 \pmod 4}} p^2.$$

Taking the logarithm

$$\log f(N) \le 2\log N \sum_{\substack{p \le N \\ p \equiv 1 \pmod 4}} 1 + 2N \sum_{\substack{p \le N \\ p \equiv 1 \pmod 4}} \frac{\log p}{p-1} + \sum_{\substack{p \le t \\ p \equiv 1 \pmod 4}} \log p + \frac{N}{2}\log 2.$$

By PNT for AP and with some computations we can see that the RHS is asymptotic to $N \log N + t$. Notice that

$$f(N) \ge \prod_{n \le N} n^2 = (N!)^2 = N^{2N} + \mathcal{O}(N),$$

combining, we get that

$$2N\log N + \mathcal{O}(N) \le \log f(N) \le N\log N + t + o(N\log N)$$

now if $t \le (1-\varepsilon)N\log N$ for all large N then the above is false for sufficiently large N, which is what we wanted.

§7.6 Example 5.10 (STEMS 2020 B3/C5, Arka Karmakar)

Clearly $b \neq 1$. Note that the leading coefficient of f must be positive and $f \in \mathbb{Q}[x]$. For now assume that f is non-constant. Consider the following claims :

> **Claim 1 –** For $n, i \in \mathbb{N}$ let $f(a^n + a^i) = b^{t(n,i)} + b^{m(n,i)}$ where $t(n,i) \ge m(n,i)$. And let $i_1, i_2, i_3, \ldots, i_k \in \mathbb{N}$, then it follows that $t(n,i_1) = t(n,i_2) = \ldots = t(n,i_k)$ for all large n.

Proof. Let $i > j$ be two positive integers. It is obvious that $t(n,i) \geq t(n,j)$ for all large n. Observe that

$$\begin{aligned}
1 = \lim_{n\to\infty} \frac{f(a^n+a^i)}{f(a^n+a^j)} &= \lim_{n\to\infty} \frac{b^{t(n,i)}+b^{m(n,i)}}{b^{t(n,j)}+b^{m(n,j)}} \\
&= \lim_{n\to\infty} \frac{1+b^{m(n,i)-t(n,i)}}{b^{t(n,j)-t(n,i)}+b^{m(n,j)-t(n,i)}} \\
&\geq \lim_{n\to\infty} \frac{1}{b^{-(t(n,i)-t(n,j))}+b^{-(t(n,i)-m(n,j))}} \\
&\geq \lim_{n\to\infty} \frac{1}{2b^{-(t(n,i)-t(n,j))}} \\
&= \frac{1}{2}\lim_{n\to\infty} b^{t(n,i)-t(n,j)}.
\end{aligned}$$

If $b > 2$ we get $t(n,i) = t(n,j)$ for all sufficiently large n. So let $b = 2$ then either $t(n,i) = t(n,j)$ for all sufficiently large n or $t(n,i) = t(n,j)+1$ for all sufficiently large n. We assume the later. Then note that,

$$1 = \lim_{n\to\infty} \frac{b^{t(n,i)}+b^{m(n,i)}}{b^{t(n,j)}+b^{m(n,j)}} = \lim_{n\to\infty} \frac{2+2^{m(n,i)-t(n,j)}}{1+2^{m(n,j)-t(n,j)}} \geq \lim_{n\to\infty} \frac{2}{1+2^{m(n,j)-t(n,j)}}$$

which implies that $m(n,j) = t(n,j)$ for all large n. Let $i > j > 1$. Hence we obtain

$$\begin{aligned}
f(a^n+a^i) &= 2^{t(n,j)+1}+2^{m(n,i)} \\
f(a^n+a^j) &= 2^{t(n,j)+1} \\
f(a^n+a) &= 2^{t(n,1)}+2^{m(n,1)}
\end{aligned}$$

for all large n. We now must have $t(n,1) = t(n,j)$. Now again using the same reasoning as above we will get $m(n,1) = t(n,1)$ which will mean $f(a^n+a^j) = f(a^n+a)$ for all large n. Contradiction! Hence the claim. □

Let us introduce some notation : Let $i \in \mathbb{N}$ and define t_n and $A(n,i)$ such that

$$f(a^n+a^i) = b^{t_n}+b^{A(n,i)}$$

for all large n (here we are using claim 1 and t_n is independent of i for small i). Let $f(x) = x^d(xg(x)+c)$ where $c \neq 0$ and $g \in \mathbb{Q}[x]$.

Claim 2 – Let p be a prime such that $p \mid b$ then $p \mid a$.

Proof. Assume that $\gcd(a,b) = 1$. Let r be some positive integer. Notice that $a^{c_1\phi(b^r)+d_1}+a^{c_2\phi(b^r)+d_2} \equiv a^{d_1}+a^{d_2} \pmod{b^r}$. Therefore if we take $c_1, c_2 \to \infty$ then using Claim 1, we get $b^r \mid f(a^{c_1\phi(b^r)+d_1}+a^{c_2\phi(b^r)+d_2}) \implies b^r \mid f(a^{d_1}+a^{d_2})$. Now taking r to be sufficiently large we get $f(a^{d_1}+a^{d_2})$ which means that $f \equiv 0$, this is a contradiction to our assumption that f is non-constant. □

Claim 3 – Let $N \in \mathbb{N}$ be a constant. Then it follows that $\{A(n,i)\}_{i=1}^{N}$ forms an A.P. for large enough n and a is a power of b.

Proof. Let $p \mid \gcd(a,b)$ be a prime. Now consider

$$(a^n+a^i)^d((a^n+a^i)g(a^n+a^i)+c) = b^{t_n}+b^{A(n,i)}$$

Taking ν_p of both sides and $n \to \infty$,

$$di\nu_p(a) + \nu_p(c) = A(n,i)\nu_p(b) \implies A(n,i) = \frac{id\nu_p(a) + \nu_p(c)}{\nu_p(b)}$$

Hence the claim. Notice that the above also gives us that $\nu_p(b)(A(n,i+1) - A(n,i)) = d\nu_p(a)$, which means both a and b have the same set of prime divisors. Now if a prime q divides both a and b then by the same reasoning we have that

$$A(n,i) = \frac{id\nu_q(a) + \nu_q(c)}{\nu_q(b)} = \frac{id\nu_p(a) + \nu_p(c)}{\nu_p(b)}$$

taking $i = 1, 2$ we get that

$$\frac{\nu_q(a)}{\nu_q(b)} = \frac{\nu_p(a)}{\nu_p(b)} \implies a = b^r$$

for some $r \in \mathbb{N}$. □

Finishing the problem is easy using the above claim.

§8 Proofs of selected Theorems

§8.1 Theorem 2.7 (Kronecker's Theorem)

Pick any $n \in \mathbb{N}$. By the pigeonhole principle, there are two multiples of k whose fractional part lie within $1/n$ of each other (to see this, divide $[0,1]$ into n equal intervals). Taking the difference, there is a multiple of k with fractional part less than $1/n$. It follows that every $x \in [0,1]$ is within $1/m$ of some $\{nk\}$, for any m. It is easy to see that we are done.

§8.2 Theorem 2.10 (Equidistribution Theorem)

(Proof by `mathcool2009`) Define $I = [a,b]$ and let S denote the set of non-negative integers n for which $\{nk\} \in I$. We want to show that $\lim_{n\to\infty} \frac{s_n}{n} = b - a$, where $s_n = |S \cap \{1, 2, \cdots, n\}|$.

By Kronecker's theorem, for any $\varepsilon > 0$ there is a positive integer N for which $\{Nk\} < \epsilon$. Now take an arbitrary integer $0 \le i < N$ and consider the infinite sequence T_i given by

$$\{ik\}, \{(i+N)k\}, \{(i+2N)k\}, \{(i+3N)k\}, \ldots$$

Assume ε is sufficiently small. Observe that there must be runs of terms in I separated with runs of terms not in I. It is easy to see that the runs of terms in I has length $\left\lfloor \frac{b-a}{L} \right\rfloor$ or $\left\lceil \frac{b-a}{L} \right\rceil$ where $L = \{Nk\}$. Similarly, the runs of terms not in I has length either $\left\lfloor \frac{1-(b-a)}{L} \right\rfloor$ or $\left\lceil \frac{1-(b-a)}{L} \right\rceil$. This means that in the limit, between

$$u_- = \left\lfloor \frac{b-a}{L} \right\rfloor \Big/ \left(\left\lfloor \frac{b-a}{L} \right\rfloor + \left\lceil \frac{1-(b-a)}{L} \right\rceil \right)$$

and

$$u_+ = \left\lceil \frac{b-a}{L} \right\rceil \Big/ \left(\left\lceil \frac{b-a}{L} \right\rceil + \left\lfloor \frac{1-(b-a)}{L} \right\rfloor \right)$$

of the terms of T_i are in I. (More precisely, let t_n denote the number of terms of T_i that are in I within the first n terms of T_i. Then $\liminf_{n\to\infty} \frac{t_n}{n} \ge u_-$ and $\limsup_{n\to\infty} \frac{t_n}{n} \le u_+$.)

Of course, as we sum over i, we see that we must have $\liminf_{n\to\infty} \frac{s_n}{n} \ge u_-$ and $\limsup_{n\to\infty} \frac{s_n}{n} \le u_+$.

Finally, since ε was arbitrary, we can choose ε small enough such that u_- and u_+ both approach $b - a$. Hence $\liminf_{n\to\infty} \frac{s_n}{n} = \limsup_{n\to\infty} \frac{s_n}{n} = b - a$, as desired.

§8.3 Theorem 3.1 (Mertens' Second Theorem)

Theorem weak form of Mertens' Second Theorem —

$$\sum_{p\le n}\frac{1}{p}=\log\log n+\mathcal{O}(1)$$

We only show the following:

$$\sum_{p\le n}\frac{1}{p}\ge\log\log n+\mathcal{O}(1)$$

The proof is motivated from the Euler product formula for the ζ function.

Demonstration.

1. Look at Digression 2.5.
2. You would want to somehow "truncate" the Euler product formula so that you get information about sum of reciprocals of primes *till* n.
3. Show that
$$\log\prod_{p\le n}\left(1-\frac{1}{p}\right)^{-1}=\sum_{p\le n}\frac{1}{p}+\mathcal{O}(1)$$
using the Taylor series of $\log(1-x)$.
4. Get a lower bound on $\prod_{p\le n}(1-\frac{1}{p})^{-1}$.
5. Conclude.

Readers interested in the complete proof may have a look at Abel Summation Formula[3] and start with Mertens' First Theorem.

§9 Acknowledgements

I am very thankful to `Superguy`[4] for valuable suggestions and for being a problem resource.

References

[1] https://artofproblemsolving.com

[2] https://math.stackexchange.com

[3] An Introduction to The Theory of Numbers, Ivan Niven, Herbert S. Zuckerman, Hugh L. Montgomery, Wiley India Pvt. Ltd.

[4] Problems from the Book, Titu Andreescu, Gabriel Dospinescu, XYZ Press

[5] Straight from the Book, Titu Andreescu, Gabriel Dospinescu, XYZ Press

[3] https://en.wikipedia.org/wiki/Abel%27s_summation_formula
[4] AoPS user : https://artofproblemsolving.com/community/user/388865