A Taste of Analytic Number Theory, Part II[1]

Ayan Nath

HS Student, Kaliabor College

Abstract. These series of articles (three in total) are 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 started with the Prime Number Theorem, giving an elementary proof of the weak version and establishing a few well known estimates for the two Chebyshev functions. We also showed Mertens' first theorem on the fly. In this part we discuss Mertens' second theorem, asymptotic density and equidistribution theorem. In the concluding part we will present some problems.

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 :

[1] **Editor's Note:** Part I (in Volume 67) contained section 1.

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 $n \leq N$ such that n^2+1 is square-free using truncated Inclusion Exclusion Principle for some fixed N.

1. Define $A_N = \{n^2+1 \mid n \leq N,\ n^2+1 \text{ is square-free}\}$. 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

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

   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

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

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 this part 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 is not 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 this part 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 know 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 part for the proof of one direction.

4. PNT for Arithmetic Progressions

This section unfortunately will 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 $a_n = a + nd$ be an arithmetic progression for relatively prime positive integers a and d. Then the number of primes in $\{a_1, a_2, \ldots, a_n\}$ is asymptotic to

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

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

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

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

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

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

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.

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:

$$\sum_{i=1}^{\infty} \frac{1}{a_n} \leq \prod_{i=1}^{k} \left(1 + \frac{1}{q_i} + \frac{1}{q_i^2} + \cdots\right).$$

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[2].

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.

[2] 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.

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, j-th 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 Section 1.2 in the last issue).
2. If $p \leq N$ is a 1 (mod 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 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. Proofs of selected Theorems

6.1. 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.

6.2. 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.

6.3. Theorem 3.1 (Mertens' Second Theorem)

We only show the following:

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

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
$$\log \prod_{p \leq n} \left(1 - \frac{1}{p}\right)^{-1} = \sum_{p \leq n} \frac{1}{p} + \mathcal{O}(1),$$
using the Taylor series of $\log(1 - x)$.
4. Get a lower bound on $\prod_{p \leq n}(1 - \frac{1}{p})^{-1}$.
5. Conclude.

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

[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