AVERAGE VALUES OF ARITHMETIC FUNCTIONS

ROSE WONG

ABSTRACT. In this paper, we will present problems involving average values of arithmetic functions. The arithmetic functions we discuss are: (1)the number of representations of natural numbers as a sum of two squares and (2)as a sum of three squares, (3)the number of decompositions of natural numbers into sums of consecutive primes, (4)the number of primitive pythagorean triangles with a hypotenuse of a given length, and (5)the number of divisors of natural numbers.

1. INTRODUCTION

An arithmetic function is defined to be a function $f(n)$, defined for $n \in \mathbb{N}$, which maps to a complex number such that f: $\mathbb{N} \to \mathbb{C}$. Examples of arithmetic functions include: the number of primes less than a given number n, the number of divisors of n, and the number of ways n can be represented as a sum of two squares. While the behavior of values of arithmetic functions are hard to predict, it is easier to analyze the behavior of the averages of arithmetic functions which we define as:

$$\lim_{n \to \infty} \frac{f(1) + ... + f(n)}{n} = L$$

where L the average value of $f(n)$. In this paper, we will examine averages of several different arithmetic functions.

2. AVERAGE NUMBER OF REPRESENTATIONS OF A NATURAL NUMBER AS A SUM OF TWO SQUARES

Let $r(n)$ be the number of representations of n as a sum of two squares

$$n = x^2 + y^2.$$

Theorem 2.1. *The average number of representations of a natural number as a sum of two squares is π. That is,*

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

Proof. For a fixed n, let

$$R(n) = r(1) + ... + r(n)$$

We interpret $r(n)$ to be the number of lattice points (points with integer coordinates) on the circle $x^2 + y^2 = n$. $R(n)$ is the number of lattice points in the disc $x^2 + y^2 \le n$ except for the origin.

If we place a unit square (centered on the lattice point with sides parallel to the coordinate axes) on each lattice point covered by the disk, the total area of these squares is $1 + R(n)$, because $R(n)$ does not include the origin.

This is not exactly the area of the disk, as some squares go beyond the disk's boundary and some areas of the circle are not covered by any square. However, we can find a circle that circumscribes all the squares and a circle that is inscribed by all the squares. These will be the bounding outer and inner circles, respectively.

The radius of the outer circle is $\sqrt{n} + \frac{1}{\sqrt{2}}$, because the greatest distance from the origin (0,0) to an outermost lattice point is $\sqrt{n}$ and the greatest distance from that lattice point to the edge of its corresponding unit square is $\frac{1}{\sqrt{2}}$.

Hence, the area of the outter circle is larger than the area of the squares.

$$1 + R(n) < \pi(\sqrt{n} + \frac{1}{\sqrt{2}})^2.$$

By a similar argument, the area of the inner circle is less than the area of the squares.

$$1 + R(n) > \pi(\sqrt{n} - \frac{1}{\sqrt{2}})^2$$

We can simplify the inequalities using the fact that $\pi\sqrt{2} < 5$ and $0 < \frac{\pi}{2} - 1 < \sqrt{n}$ ($n = 1, 2, 3, ...$). Therefore,

$$R(n) < \pi(\sqrt{n} + \frac{1}{\sqrt{2}})^2 - 1 = \pi n + \pi\sqrt{2}\sqrt{n} + \frac{\pi}{2} - 1 < \pi n + 6\sqrt{n}$$

and

$$R(n) > \pi(\sqrt{n} - \frac{1}{\sqrt{2}})^2 - 1 = \pi n - \pi\sqrt{2}\sqrt{n} + \frac{\pi}{2} - 1 > \pi n - 6\sqrt{n}.$$

It follows that

$$\pi n - 6\sqrt{n} < R(n) < \pi n + 6\sqrt{n},$$

which is the same as

$$|R(n) - \pi n| < 6\sqrt{n}.$$

If we divide by n, we get

$$|\frac{R(n)}{n} - \pi| < \frac{6}{\sqrt{n}}.$$

Taking the limit as $n \to \infty$, we have $R(n)/n \to \pi$ and the proof is complete. $\square$

3. Average Number of Representations of a Natural Number as a sum of three squares

We will now extend the previous problem, that is, we will find the average number of representations of a natural number as a sum of three squares. Again, we will use geometry with integer lattice points to approach this problem.

Let $f(n)$ denote the number of integral solutions of

$$x^2 + y^2 + z^2 = n.$$

Theorem 3.1. *The average number of representations of a natural number as a sum of three squares is $\frac{4}{3}\pi$.*

$$\lim_{n \to \infty} \frac{f(1) + f(2) + ... + f(n)}{n^{3/2}} = \frac{4}{3}\pi$$

Proof. For a fixed n, let

$$F(n) = f(1) + ... + f(n).$$

We interpret $f(n)$ to be the number of lattice points on the sphere $x^2 + y^2 + z^2 = n$, so $F(n)$ to be the number of lattice points in the sphere $x^2 + y^2 + z^2 \leq n$ except for the origin. If we place a unit cube on each lattice point covered in the sphere, the total volume of these cubes is $1 + F(n)$.

We want to find a sphere that encloses all of the cubes and a sphere that is inscribed inside the cubes. The radius of the outer sphere is $\sqrt{n} + \frac{\sqrt{3}}{2}$. Since the greatest distance between outermost lattice point and the edge of its corresponding cube is $\sqrt{(\frac{1}{2})^2 + (\frac{1}{2})^2 + (\frac{1}{2})^2} = \frac{\sqrt{3}}{2}$. Similarly, the radius of the inner sphere is $\sqrt{n} - \frac{\sqrt{3}}{2}$.

Comparing the volumes, we have

$$1 + R(n) < \frac{4}{3}\pi(\sqrt{n} + \frac{\sqrt{3}}{2})^3$$

and

$$1 + R(n) > \frac{4}{3}\pi(\sqrt{n} - \frac{\sqrt{3}}{2})^3.$$

Expanding the inequalities out, we have

$$R(n) < (\frac{4}{3}\pi n^{3/2} + 3\pi\sqrt{n} - 1) + (2\sqrt{3}\pi n + \frac{\sqrt{3}}{2}\pi)$$

and

$$R(n) > (\frac{4}{3}\pi n^{3/2} + 3\pi\sqrt{n} - 1) - (2\sqrt{3}\pi n + \frac{\sqrt{3}}{2}\pi).$$

Using the fact that $2\sqrt{3}\pi < 11$ and $\frac{\sqrt{3}}{2}\pi < 3$, we have

$$|R(n) - \frac{4}{3}\pi n^{3/2} - 3\pi\sqrt{n} + 1)| < 11n - 3.$$

Dividing by $n^{3/2}$ gives

$$|\frac{R(n)}{n^{3/2}} - \frac{4}{3}\pi - \frac{3\pi}{n} + \frac{1}{n^{3/2}})| < \frac{11}{\sqrt{n}} - \frac{3}{n^{3/2}}.$$

Taking the limit as $n \to \infty$, we have $\frac{R(n)}{n^{3/2}} \to \frac{4}{3}\pi$ and the proof is complete. $\square$

4. AVERAGE VALUE OF NUMBER OF DECOMPOSITIONS OF A NATURAL NUMBER n INTO A SUM OF CONSECUTIVE PRIMES

Let $f(n)$ be the number of decompositions of a natural number n into a sum of one or more consecutive prime numbers. For example, $f(395) = 2$ because

$$395 = 127 + 131 + 137 = 71 + 73 + 79 + 83 + 89.$$

Theorem 4.1. *The average number of decompositions of a natural number into a sum of one or more consecutive prime numbers is* $\log 2$. *[1]*

$$\lim_{n\to\infty} \frac{f(1) + ... + f(n)}{n} = \log 2$$

Proof. Let P be a sequence of consecutive primes: $p_1 < p_2 < p_3 < ...$ Every set of consecutive primes whose sum is less than or equal to x contributes 1 to the sum $f(1) + f(2) + ... + f(n)$. The number of sets of r primes that satisfy the condition is at most $\pi(\frac{x}{r})$ and at least $\pi(\frac{x}{r}) - r$.

Therefore, we have

$$\sum_{r=1}^{k}(\pi(\frac{x}{r}) - r) \le f(1) + f(2) + ... + f(x) \le \sum_{r=1}^{k}(\pi(\frac{x}{r})),$$

where k is determined by

$$p_1 + p_2 + ... + p_k \leq x < p_1 + p_2 + ... + p_{k+1}$$

Using the above inequality and $p_r \asymp r \log r$ we get

$$k \asymp \sqrt{\frac{x}{\log x}}$$

This implies that $\sum_{r=1}^{k} r = O(x)$. If we can show that $\sum_{r=1}^{k} (\pi(\frac{x}{r})) \sim x \log 2$, it will follow that $f(1) + f(2) + ... + f(x) \sim x \log 2$

Using the above relationship and the prime number theorem, we have

$$\sum_{r=1}^{k} (\pi(\frac{x}{r})) \sim \sum_{r=1}^{k} (\frac{x}{r \log x/r}) \sim \int_1^k \frac{x}{r \log x/r} \, dr = [-x \log(\log x/r)]_1^k$$

$$= x(\log \log x - \log \log \frac{x}{k}) \sim x(\log \log x - \log \log \sqrt{x \log x}) \sim x \log 2.$$

□

Notation explanation: Let f and g be functions of x. The notation $f \asymp g$ denotes that $f(x)/g(x)$ is bounded above and below by positive numbers for large values of x. The notation $f = O(g)$ denotes that $\exists c$ such that $|f(x)| \leq cg(x)$. The notation $f \sim g$ denotes that $\lim_{x \to \infty}, \frac{f(x)}{g(x)} = 1$

5. AVERAGE VALUE OF PRIMITIVE PYTHAGOREAN TRIANGLES WITH HYPOTENUSE EQUAL TO N

Let $P(n)$ be the number of primitive Pythagorean triangles with hypotenuse equal to n. For example, P(5)=1 and and P(65)=2 because

$$3^2 + 4^2 = 5^2,$$

$$33^2 + 56^2 = 63^2 + 16^2 = 65^2$$

We would like to determine whether the average value of $P(n)$ converges, and if so, to what value. That is,

$$\lim_{n \to \infty} \frac{P(1) + ... + P(n)}{n} = K$$

Rather than approaching this problem through mathematical arguments, we will use computer computations to infer a value. For each value of n from 1 to 1000, we generate and graph the average values up to and including n. Looking at Figure 1, we see that the average values (measured on the vertical axis) oscillate, but as $n \to \infty$, the
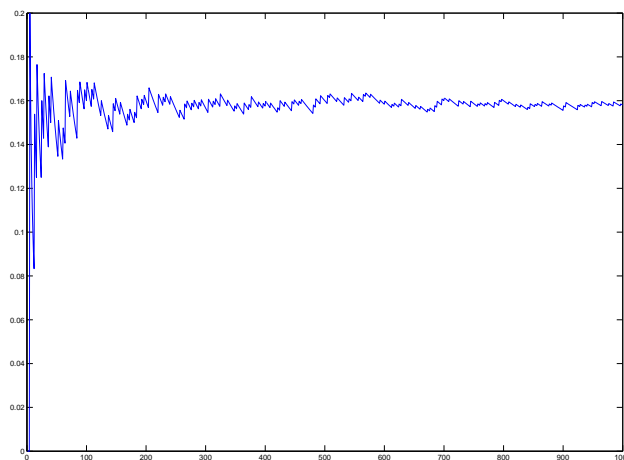
FIGURE 1. Average P(n) for n=[1,1000]

average values level off. We can conjecture that the limit as $n \to \infty$ of the average value is:

$$\lim_{n \to \infty} \frac{P(1) + ... + P(n)}{n} = 0.159.$$

6. AVERAGE VALUE OF DIVISORS OF A NATURAL NUMBERS

Let $d(n)$ be the number of divisors of the natural number n. We will show that

Theorem 6.1. *The average value of the number of divisors of natural numbers grows like* $\log n$.

$$\frac{d(1) + d(2) + ... + d(n)}{n} \sim \log n$$

Proof. Let k be a fixed integer. If we list the multiples of k less than or equal to n:

$$k, 2k, 3k, ...[\frac{n}{k}]k,$$

we find that there are $[\frac{n}{k}]$ multiples, where $[\ ]$ denotes the floor function. Each of those multiples contributes 1 to the sum $d(1) + ... + d(n)$.

If we examine multiples of all integers $k \leq n$, it follows that summing over k gives

$$\sum_{k=1}^{n}[\frac{n}{k}] = d(1) + ... + d(n)$$

Now, we want to prove that

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n}[\frac{n}{k}]}{n\log n} = 1$$

First, we establish the relationship:

$$\frac{n}{k} - 1 < [\frac{n}{k}] \leq \frac{n}{k}.$$

Summing over k gives:

$$\sum_{k=1}^{n}(\frac{n}{k} - 1) < \sum_{k=1}^{n}[\frac{n}{k}] \leq \sum_{k=1}^{n}\frac{n}{k}.$$

We then factor out n to get:

$$n\sum_{k=1}^{n}(\frac{1}{k} - \frac{1}{n}) < \sum_{k=1}^{n}[\frac{n}{k}] \leq n\sum_{k=1}^{n}\frac{1}{k}$$

The first and last term in the above inequality can be rewritten as the integrals $n\int_1^n(\frac{1}{k} - \frac{1}{n})\,dk$ and $n\int_1^n\frac{1}{k}\,dk$.

Integrating gives

$$n\log n - n + 1 < \sum_{k=1}^{n}[\frac{n}{k}] \leq n\log n.$$

So taking $n \to \infty$, we have

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n}[\frac{n}{k}]}{n\log n} = 1.$$

And so,

$$\frac{d(1) + d(2) + ... + d(n)}{n} \sim \log n.$$

$\square$

REFERENCES

[1] Moser, L., "On the Sum of Consecutive Primes," Notes on Number Theory III, Canadian Mathematical Bulletin 6, p 159-161, 1963.
[2] Young, Robert M., "Excursions in Calculus," The Mathematical Association of America, USA, 1992.