COMPUTING SUMS AND THE AVERAGE VALUE OF THE DIVISOR FUNCTION

Abstract. We introduce a method for computing sums of the form

$$\sum_{n \le x} f(n)$$

where $f(n)$ is "nice". We apply this method to study the average value of $d(n)$, where $d(n)$ is the number of positive divisors of n.

1. Introduction: Computing sums

Let $x \ge 1$ be an integer. Consider the problem of summing

$$1 + 2 + 3 + \cdots + (x-1) + x = \sum_{n=1}^{x} n = \sum_{n \le x} n.$$

We will compute this sum in two different ways. First, we observe that if

$$S = \sum_{n \le x} n,$$

then

$$\begin{aligned} S &= 1 + 2 + 3 + 4 + \cdots + (x-1) + x \\ S &= x + (x-1) + (x-2) + (x-3) + \cdots + 2 + 1. \end{aligned}$$

Summing the left hand side and the right hand side, we have that

$$2S = x(x+1).$$

Solving for S, we have that

$$S = \sum_{n \le x} n = \frac{x(x+1)}{2} = \frac{x^2}{2} + \frac{x}{2}.$$

As an example,

$$1 + 2 + 3 + 4 + 5 + 6 + 7 + \cdots + 99 + 100 = \sum_{i=1}^{100} i = \frac{100 \times 101}{2} = 5050.$$

To describe the second method of computing S, we will use the fact that

$$S = \sum_{n \le x} n$$

Date: March 21, 2017.

is approximately equal to the area under the curve $f(n) = n$ from $n = 1$ to $n = x$. Note that this area, which we write as

$$\int_1^x t \, \mathrm{d}t,$$

is equal to

$$\frac{x^2}{2} - \frac{1}{2}.$$

(Think of the area under $f(n) = n$ as a right isosceles triangle with the two equal sides of length x.) We are almost there; we are just missing $\frac{x}{2} + \frac{1}{2}$. To recover this last bit, we introduce the **floor function**

$$\lfloor t \rfloor = \text{largest integer } n \text{ such that } n \leq t$$

For example,

$$\lfloor 1 \rfloor = 1, \quad \lfloor 2.999 \rfloor = 2, \quad \lfloor \pi \rfloor = 3.$$

(We sometimes call $\lfloor x \rfloor$ the **integer part of** x.) If we write A for the area under the curve $f(t) = t$ from $t = 1$ to $t = x$, then we find that

$$S - A = 1 + \int_1^x t \, \mathrm{d}t - \int_1^x \lfloor t \rfloor \, \mathrm{d}t = 1 + \int_1^x (t - \lfloor t \rfloor) \, \mathrm{d}t.$$

That is, $S - A$ equals the area under the curve $g(t) = t - \lfloor t \rfloor$ from $t = 1$ up to $t = x$. Now,

$$\int_1^x (t - \lfloor t \rfloor) \, \mathrm{d}t = 1 + \sum_{i=1}^{x-1} \int_i^{i+1} (t - i) \, \mathrm{d}t = 1 + \sum_{i=1}^{x-1} \frac{1}{2} = 1 + \frac{x-1}{2} = \frac{x}{2} + \frac{1}{2},$$

which is the remaining piece.

This technique of relating a sum to an integral (i.e., area under a curve) is a very important technique in modern mathematics because we have much better knowledge of how to compute integrals than we do sums.

Theorem 1.1 (Euler's summation formula)**.** *Suppose that $f(t)$ is a function whose derivative $f'(t)$ is continuous on the interval $1 \leq t \leq x$, where x is an integer. Then*

$$f(1) + f(2) + f(3) + \cdots + f(x-1) + f(x) = \sum_{n \leq x} f(n)$$

equals

$$f(1) + \int_1^x f(t) \, \mathrm{d}t + \int_1^x (t - \lfloor t \rfloor) f'(t) \, \mathrm{d}t.$$

Recall that $f'(t)$, the *derivative of* f, is the *slope* of f at the point $(t, f(t))$.

Exercise 1.2. Show that

$$1^2 + 2^2 + 3^2 + 4^2 + \cdots + (x-1)^2 + x^2 = \sum_{n \leq x} n^2 = \frac{x^3}{3} + \frac{x^2}{2} + \frac{x}{6}$$

for any integer $x \geq 1$. (Hint: If $f(t) = t^2$, then $\int_1^x t^2 \, \mathrm{d}t = (x^3 - 1)/3$ and $f'(t) = 2t$ on the interval $1 \leq t \leq x$.)

Exercise 1.3. Show that

$$1^3 + 2^3 + 3^3 + 4^3 + \cdots + (x-1)^3 + x^3 = \sum_{n=1}^{x} n^3 = \frac{x^4}{4} + \frac{x^3}{2} + \frac{x^2}{4}$$

for any integer $x \geq 1$. (Hint: If $f(t) = t^3$, then $\int_1^x t^3 \text{ d}t = (x^4 - 1)/4$ and $f'(t) = 3t^2$ on the interval $1 \leq t \leq x$.)

2. A important example: The harmonic series

One of the most important sums in mathematics is the **harmonic sum**

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \cdots + \frac{1}{x-1} + \frac{1}{x} = \sum_{n=1}^{x} \frac{1}{n}.$$

(Remember, we take x to be a positive integer here.) We can now compute this sum *exactly* using Theorem 1.1 and the fact that

$$\int_1^x \frac{1}{t} dt = \ln x,$$

where $\ln x$ is the natural logarithm of x. Recall that

$$\log_{10}(x) = \frac{\ln x}{\ln 10}, \qquad \ln 10 = 2.30258509299405\ldots$$

Exercise 2.1. Show that if $n \geq 2$ is an integer and $\hat{n}$ is the number of digits in n, then

$$\left| \hat{n} - \frac{\ln x}{\ln 10} \right| \leq 1.$$

Using Theorem 1.1, we have that

$$\sum_{n \leq x} \frac{1}{n} = 1 + \int_1^x \frac{1}{t} dt - \int_1^x \frac{t - \lfloor t \rfloor}{t^2} dt = \ln x + 1 - \int_1^x \frac{t - \lfloor t \rfloor}{t^2} dt. \tag{2.1}$$

Exercise 2.2. Let $x \geq 2$ be an integer. Using the fact that

$$\int_1^x \frac{1}{t^2} dt = 1 - \frac{1}{x},$$

show that

$$\frac{1}{x} < 1 - \int_1^x \frac{t - \lfloor t \rfloor}{t^2} dt < 1.$$

(Hint: Show that $0 \leq t - \lfloor t \rfloor \leq 1$.)

Notation 2.3. If $g(x) > 0$ for all $x \geq a$, we say that

$$f(x) = O(g(x)) \qquad \text{(read: ``}f(x)\text{ is big oh of }g(x)\text{'')}$$

to mean that $|f(x)/g(x)|$ is bounded for $x \geq a$. That is, there exists a constant $M > 0$ such that

$$|f(x)| \leq M \cdot g(x) \text{ for all } x \geq a.$$

An equation of the form

$$f(x) = h(x) + O(g(x))$$

means that $f(x) - h(x) = O(g(x))$. If $f(t) = O(g(t))$ for $t \geq a$, then $\int_a^x f(t) \text{ d}t = O(\int_a^x g(t) \text{ d}t)$ for $x \geq a$.

Exercise 2.4. Prove the following:

- $\sin(x) = O(1)$ for $x \geq 0$.
- $\log x = O(x)$ for $x \geq 1$.
- $\sqrt{x+1} = O(\sqrt{x})$ for $x \geq 1$.
- $\lfloor x \rfloor = x + O(1)$ for $x \geq 0$.

Exercise 2.2 shows us that

$$\frac{1}{x} < \sum_{n \leq x} \frac{1}{n} - \ln x < 1.$$

for all integers $x \geq 2$. Thus

$$\sum_{n \leq x} \frac{1}{n}$$

is very-well approximated by $\ln x$ for all values of $x \geq 2$, but for *all* integers $x \geq 2$,

$$\sum_{n \leq x} \frac{1}{n} \neq \ln x.$$

Exercise 2.5. Show that if $x \geq 1$, then

$$1 - \int_1^x \frac{t - \lfloor t \rfloor}{t^2} dt = \gamma + O\Big(\frac{1}{x}\Big),$$

where $\gamma = 0.577215664901\ldots$ Thus

$$\sum_{n \leq x} \frac{1}{n} = \log x + \gamma + O\Big(\frac{1}{x}\Big)$$

for all $x \geq 1$.

Exercise 2.6. Show that as x grows, the quantity

$$\sum_{n \leq x} \frac{1}{n} - \ln x$$

approaches γ.

Exercise 2.7. Determine whether γ can be expressed as the ratio of two integers a/b, where $a \neq b \neq 0$ and a has no prime factors in common with b.

3. The average behavior of the divisor function

Let $n \geq 2$ be a positive integer. In the study of numbers, it is very important to understand the number of positive integers that divide n. To this end, we define

$$\tau(n) = \text{ the number of distinct positive integers } d \text{ such that } d \text{ divides } n.$$

For example, $\tau(1) = 1$, $\tau(2) = 2$ because 1 and 2 divide 2, $\tau(3) = 2$ because 1 and 3 divide 3, $\tau(4) = 3$ because 1 and 2 and 4 divide 4, and so on. Let's explore the behavior of $\tau(n)$.

Exercise 3.1. Compute $\tau(n)$ for all integers $1 \leq n \leq 100$. Try to find some patterns.

Exercise 3.2. Suppose that p is prime. What is $\tau(p^2)$? $\tau(p^3)$? What is $\tau(p^k)$ for any positive integer $k \geq 1$?

Let $p_1 = 2$, $p_2 = 3$, $p_3 = 5$, $p_4 = 7$, and in general, let p_n denote the n-th prime. Define

$$n\# = p_1 p_2 p_3 \cdots p_{n-1} p_n.$$

Exercise 3.3. What is $\tau(3\#)$? $\tau(4\#)$? $\tau(5\#)$? How large is $\tau(n\#)$ for any integer $n \geq 1$?

The preceding exercises show that $\tau(n)$ behaves pretty wildly. Notice that $\tau(p) = 2$ for every prime p, but $\tau(n\#)$ grows very quickly. This begs the question:

How do we study $\tau(n)$?

We can take a cue from statistics and study the **mean value** of $\tau(n)$. That is, we study

$$\frac{1}{x} \sum_{n \leq x} \tau(n)$$

instead of $\tau(n)$.

We now introduce the notation $d \mid n$ to say that d divides n. One way that we can write $\tau(n)$ is

$$\tau(n) = \sum_{d \mid n} 1.$$

That is, we sum 1 for every distinct divisor d of n. Now,

$$\frac{1}{x} \sum_{n \leq x} \tau(n) = \frac{1}{x} \sum_{n \leq x} \sum_{d \mid n} 1.$$

We now observe that *divisors come in pairs*: if $d \mid n$, then we can write $n = dq$ for some integer $q \geq 1$. (This might be the same as d if, say, $n = d^2$.) Thus we can rewrite the sum as

$$\frac{1}{x} \sum_{n \leq x} \tau(n) = \frac{1}{x} \sum_{\substack{d,q \\ dq \leq x}} 1. \tag{3.1}$$

This can be interpreted as a sum extended over points with integer coordinates in the dq-plane; we call such points **lattice points**. The lattice points in $dq = n$ lie on a hyperbola, so the inner sum in (3.1) counts the number of lattice points which lie on the hyperbolas $dq = 1$, $dq = 2$, $dq = 3, \ldots, dq = \lfloor x \rfloor$.

For each fixed $d \leq x$, we can count first those lattice points on the horizontal line segment $1 \leq q \leq x/d$, and then sum over all $d \leq x$. Therefore,

$$\frac{1}{x} \sum_{n \leq x} \tau(n) = \frac{1}{x} \sum_{d \leq x} \sum_{q \leq x/d} 1. \tag{3.2}$$

Now,

$$\sum_{q \leq x/d} 1 = \left\lfloor \frac{x}{d} \right\rfloor = \frac{x}{d} + O(1),$$

so

$$\sum_{d \leq x} \sum_{q \leq x/d} 1 = \sum_{d \leq x} \left(\frac{x}{d} + O(1) \right) = x \sum_{d \leq x} \frac{1}{d} + O\Big(\sum_{d \leq x} 1 \Big) = x \sum_{d \leq x} \frac{1}{d} + O(x).$$

But we already proved that

$$\sum_{d\le x}\frac{1}{d}=\log x+\gamma+O\Big(\frac{1}{x}\Big).$$

Thus

$$\sum_{n\le x}\tau(n)=\sum_{d\le x}\sum_{q\le x/d}1=x\ln x+O(x),$$

so

$$\frac{1}{x}\sum_{n\le x}\tau(n)=\ln x+O(1).$$

We conclude

Theorem 3.4. *The mean value of* $\tau(n)$ *is* $\ln n$.

Exercise 3.5. Show that for $x\ge 1$,

$$\begin{aligned}\sum_{n\le x}\ln n&=\int_1^x \ln t\ \mathrm{d}t+\int_1^x\frac{t-\lfloor t\rfloor}{t}\ \mathrm{d}t\\&=x\log x-x+1+\int_1^x\frac{t-\lfloor t\rfloor}{t}\ \mathrm{d}t\\&=x\log x-x+O(\log x).\end{aligned}$$

(Hint: Use the fact that $\int_1^x \ln t\ \mathrm{d}t=x\ln x-x+1$ and $(\ln t)'=1/t$.)

Exercise 3.6. Show that as x grows,

$$\frac{1}{x}\sum_{n\le x}(\tau(n)-\ln n)$$

approaches zero.

Define

$$x!=1\times 2\times 3\times 4\times\cdots\times(x-1)\times x.$$

Exercise 3.7. Show that $\ln(x!)=x\ln x-x+O(\ln x)$.

Exercise 3.8. Show that

$$\ln(x!)=x\ln x-x-\frac{1}{2}\ln x+\frac{\ln 2\pi}{2}+O\Big(\frac{1}{x}\Big).$$

So, as x grows, $x!$ approaches

$$\sqrt{\frac{2\pi}{x}}\Big(\frac{x}{e}\Big)^x.$$