The average order of the divisor function

The divisor function $\tau(n)$ gives the number of (positive) divisors of the natural number $n$. Locally it appears to be a bit erratic, but we can smooth it out by averaging. In fact it is easy to see that

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

(1)

Here’s the one-line proof: We have

$$\sum_{n \leq x} \tau(n) = \sum_{n \leq x} \sum_{d|n} 1 = \sum_{d \leq x} \sum_{n \leq x \atop d|n} 1 = \sum_{d \leq x} \left\lfloor \frac{x}{d} \right\rfloor = \sum_{d \leq x} \left( \frac{x}{d} + O(1) \right) = x \log x + O(x).$$

The error here is not caused by the sum of $1/d$, which we know to fairly good precision, but from the fact that we are making an error of $O(1)$ for each term $d \leq x$, which adds up to something significant.

This result can be improved significantly by realizing that the divisors of $n$ essentially come in pairs, where each divisor $d$ of $n$ with $d < \sqrt{n}$ is paired with the divisor $n/d > \sqrt{n}$. The only divisor which might not be paired up is $\sqrt{n}$, which is a divisor precisely when $n$ is a square. (This proves a book exercise in Ch. 2 that $\tau(n)$ is odd if and only if $n$ is a square.) Using this thought and that $d < \sqrt{n}$ if and only if $d < n/d$, we have

$$\tau(n) = \begin{cases} 
2 \sum_{d|n \atop d<n/d} 1, & n \text{ is not a square,} \\
1 + 2 \sum_{d|n \atop d<n/d} 1, & n \text{ is a square.}
\end{cases}$$

Thus,

$$\sum_{n \leq x} \tau(n) = \sum_{n \leq x} 2 \sum_{d|n \atop d<n/d} 1 + \sum_{n^2 \leq x} 1$$

$$= 2 \sum_{d<\sqrt{x}} \sum_{n \leq x \atop d|n \atop n/d>d} 1 + \left\lfloor \sqrt{x} \right\rfloor$$

$$= 2 \sum_{d<\sqrt{x}} \left( \left\lfloor \frac{x}{d} \right\rfloor - d \right) + \left\lfloor \sqrt{x} \right\rfloor.$$
Do you see the transition from the second line to the third? We are counting multiples of \(d\), say the numbers \(jd\), where \(d < j \leq x/d\). The number of integers \(j\) in this interval is \([x/d] - d\). The advantage of this approach is that the floor function now appears on about \(\sqrt{x}\) terms so that there is a much smaller error when we remove it. Taking up the calculation, we have

\[
\sum_{n \leq x} \tau(n) = 2 \sum_{d < \sqrt{x}} \left( \frac{x}{d} + O(1) - d \right) + \lfloor \sqrt{x} \rfloor = 2x \sum_{d < \sqrt{x}} \frac{1}{d} - 2 \sum_{d < \sqrt{x}} d + O(\sqrt{x}).
\]

For the first sum here we use Theorem 1 from the first week’s notes, so that

\[
2x \sum_{d < \sqrt{x}} \frac{1}{d} = 2x \left( \log \sqrt{x} + \gamma + O\left( \frac{1}{\sqrt{x}} \right) \right) = x \log x + 2\gamma x + O(\sqrt{x}).
\]

For the sum of \(d\), there’s a formula for summing consecutive integers:

\[
\sum_{j=1}^{N} = \frac{1}{2} (N + 1)N,
\]

for positive integers \(N\), which is easily proved by induction or “holistically” (it’s the number of terms times the average term). Thus,

\[
2 \sum_{d < \sqrt{x}} d = (\sqrt{x} + O(1)) \left( \sqrt{x} + O(1) \right) = x + O(\sqrt{x}).
\]

Putting these calculations into our last equation for the sum of \(\tau(n)\) and dividing by \(x\) to get the average, we have

\[
\frac{1}{x} \sum_{n \leq x} \tau(n) = \log x + 2\gamma - 1 + O\left( \frac{1}{\sqrt{x}} \right).
\]  

This is a big improvement over (1). The true order of the error term here is an unsolved problem, known as the Dirichlet divisor problem. It can be viewed geometrically as the problem of accurately counting the number of lattice points in the first quadrant of the \(u, v\) plane that are under the hyperbola \(uv = x\).

2 The average order of the sum-of-divisors function

Instead of counting divisors of \(n\) one can sum them. This function has an ancient history, first defined by Pythagoras (actually the sum of the “proper” divisors of \(n\)), and it may be the very first function defined in mathematics. Let \(\sigma(n)\) be the sum of all of the (positive) divisors of \(n\), so that Pythagoras was considering \(\sigma(n) - n\). The function \(\sigma(n)\) is somewhat easier to deal with than the more historical version, and that’s what we’ll deal with in these notes.
Like $\tau(n)$, $\sigma(n)$ appears a bit erratic locally, and it makes sense to try and smooth it out by averaging. We try the same initial technique as in the last section, getting
\[
\sum_{n\leq x} \sigma(n) = \sum_{n\leq x} \sum_{d|n} d = \sum_{d\leq x} \sum_{n\leq x/d} 1 = \sum_{d\leq x} \frac{x}{d} \left\lfloor \frac{x}{d} \right\rfloor = \sum_{d\leq x} \left( \frac{x^2}{d} + O(d) \right) = x^2 + O(x^2).
\]
This is not too cool since the error term is of the same order of magnitude as the main term. Saying that something is $x^2 + O(x^2)$ does not even assert that the something is positive much less tends to infinity. It is the same as saying that it is $O(x^2)$, which gives an upper bound on how fast the absolute value of the quantity might grow.

To handle this difficulty, we could use the same technique as in the previous section, and that’s exactly the way the book deals with it (in Ch. 3). It is easier to use another technique, one that reviews a skill from week 1. Instead of averaging $\sigma(n)$, let’s try averaging $\sigma(n)/n$. We have
\[
\frac{\sigma(n)}{n} = \frac{1}{n} \sum_{d|n} d = \frac{1}{n} \sum_{d|n} \frac{n}{d} = \sum_{d|n} \frac{1}{d},
\]
since as $d$ ranges over all divisors of $n$, so does $n/d$. Thus,
\[
\sum_{n\leq x} \frac{\sigma(n)}{n} = \sum_{n\leq x} \sum_{d|n} \frac{1}{d} = \sum_{d\leq x} \sum_{n\leq x/d} 1 = \sum_{d\leq x} \frac{1}{d} \left\lfloor \frac{x}{d} \right\rfloor = \sum_{d\leq x} \left( \frac{x}{d^2} + O\left(\frac{1}{d}\right) \right).
\]
The sum of the $O$-terms is $O(\log x)$, but how should we deal with the sum of $1/d^2$? Well, from Euler we know that the infinite sum converges to $\pi^2/6$:
\[
\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}. \tag{3}
\]
But we have a partial sum of this infinite sum. The partial sums are supposed to be converging to the infinite sum, so we write
\[
\sum_{d\leq x} \frac{1}{d^2} = \frac{\pi^2}{6} - \sum_{d>x} \frac{1}{d^2},
\]
so now we have the problem of estimating the “tail”, which involves numbers $d > x$. This can be done with an integral. It’s easy to see that
\[
\frac{1}{d^2} < \int_{d-1}^{d} \frac{1}{t^2} dt
\]
for $d \geq 2$, since the integrand is larger than $1/d^2$ and the length of the interval of integration is 1. Summing this for $d > N$, we have
\[
\sum_{d>N} \frac{1}{d^2} < \int_{N}^{\infty} \frac{1}{t^2} dt = -\frac{1}{t} \bigg|_{N}^{\infty} = \frac{1}{N}.
\]
We thus have that
\[ \sum_{d \leq x} \frac{1}{d^2} = \frac{\pi^2}{6} + O\left(\frac{1}{x}\right), \]
and using this in the above calculation for the sum of \( \sigma(n)/n \), we have
\[ \sum_{n \leq x} \frac{\sigma(n)}{n} = \frac{\pi^2}{6} x + O(\log x). \] (4)

So, we have solved another problem fairly easily. Note that we could have used the idea of small divisors of \( n \) governing the large divisors of \( n \) as we did in the previous section with \( \tau(n) \), but it wouldn’t have mattered much since the error corresponding to \( d \) is no longer \( O(1) \) but is \( O(1/d) \), so the big divisors are not so painful to deal with.

We are left though with the original problem of finding the average order of \( \sigma(n) \). But for this we can use partial summation applied to (4). We have
\[ \sum_{n \leq x} \sigma(n) = \sum_{n \leq x} \frac{\sigma(n)}{n} n = x \sum_{n \leq x} \frac{\sigma(n)}{n} - \int_{1}^{x} \sum_{n \leq t} \frac{\sigma(n)}{n} dt \]
\[ = x \left( \frac{\pi^2}{6} x + O(\log x) \right) - \int_{1}^{x} \left( \frac{\pi^2}{6} t + O(\log t) \right) dt. \]
The integral here can be evaluated, where the main term is
\[ -\frac{\pi^2}{12} t^2 \bigg|_{1}^{x} = -\frac{\pi^2}{12} x^2 + O(1), \]
and the error term is \( O(x \log x) \), using that the \( O \)-term in the integral is \( O(\log x) \), and the length of the interval is smaller than \( x \). Putting this together and dividing by \( x \), we have
\[ \frac{1}{x} \sum_{n \leq x} \sigma(n) = \frac{\pi^2}{6} x + O(\log x) - \frac{\pi^2}{12} x + O \log x = \frac{\pi^2}{12} x + O(\log x). \] (5)

So, what we’ve learned here is that sometimes a different sum is easier to do than the one you start with, and a transition can be made via partial summation. We also learned that when there is a convergent infinite series and you have a partial sum of it, you can write it as the infinite sum minus the tail, which can be estimated as part of an error term.

3 Euler’s function \( \varphi \)

Let \( \varphi \) denote Euler’s function, so that \( \varphi(n) \) is the number of integers in \([1, n]\) that are coprime to \( n \). It is the order of the unit group of the ring \( \mathbb{Z}/n\mathbb{Z} \), and so we have Euler’s theorem:
\[ a^{\varphi(n)} \equiv 1 \pmod{n} \text{ when } \gcd(a, n) = 1. \]
We would like to find its average order, and to do so it would be nice to be able to write
\[ \varphi(n) = \sum_{d|n} g(d) \]
for some arithmetic function \( d \). After all, doing this is exactly what led us to be able to compute the average order of \( \tau \) and of \( \sigma \) (and of the function \( \sigma(n)/n \)).

To solve this problem, let’s see how to compute \( \varphi(n) \) for an interesting numerical value of \( n \), namely \( n = 900 \). We are counting the number of integers in \([1, 900]\) coprime to 900, which is the same as counting the number of integers in \([1, 900]\) that are not divisible by any of 2, 3, and 5, since these are the only prime divisors of 900. This can be done by an inclusion-exclusion argument, so that
\[ \varphi(900) = 900 - \frac{900}{2} - \frac{900}{3} - \frac{900}{5} + \frac{900}{6} + \frac{900}{10} + \frac{900}{15} - \frac{900}{30} = 240. \]
(If this is bewildering, you might want to review inclusion-exclusion in a discrete math book.) We see here a sum over some of the divisors of 900, the denominators are the “squarefree” divisors of 900. (A squarefree number \( m \) is one which is not divisible by the square of any number larger than 1; it is categorized by its prime factorization, which has all different primes with no repeats.) Further, we have plus and minus signs above depending on whether the denominator \( d \) has an even or odd number of prime factors.

Let
\[ \mu(n) = \begin{cases} (-1)^k, & n \text{ is squarefree and has exactly } k \text{ prime divisors}, \\ 0, & n \text{ is not squarefree}. \end{cases} \]

Thus, the above inclusion-exclusion expression for \( \varphi(900) \) could be written as
\[ \varphi(900) = \sum_{d|900} \mu(d) \frac{900}{d}. \]

Further, there’s nothing special about “900”, the above formula works in general:
\[ \varphi(n) = \sum_{d|n} \mu(d) \frac{n}{d}. \]

With this we can fairly easily compute the average order of \( \varphi(n)/n \). We have
\[ \sum_{n \leq x} \frac{\varphi(n)}{n} = \sum_{n \leq x} \sum_{d|n} \frac{\mu(d)}{d} = \sum_{d \leq x} \frac{\mu(d)}{d} \left\lfloor \frac{x}{d} \right\rfloor = \sum_{d \leq x} \frac{\mu(d)}{d} \left( \frac{x}{d} + O(1) \right). \]
The error term is dealt with as follows:
\[ \left| \sum_{d \leq x} \frac{\mu(d)}{d} \cdot O(1) \right| = O \left( \sum_{d \leq x} \left| \frac{\mu(d)}{d} \right| \right) = O \left( \sum_{d \leq x} \frac{1}{d} \right) = O(\log x). \]
(The point is that by the triangle inequality, the terms must be made all non-negative, unless there’s finer information to be used, and we will see that this error of $O(\log x)$ is perfectly acceptable, so finer information is necessary.) The main term is

$$\sum_{d \leq x} \frac{\mu(d)}{d^2} = \sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} - \sum_{d>x} \frac{\mu(d)}{d^2}.$$

The infinite sum is some constant $\beta$, in fact the series converges absolutely. The tail sum is estimated as follows:

$$x \left| \sum_{d>x} \frac{\mu(d)}{d^2} \right| \leq x \sum_{d>x} \frac{1}{d^2} = O(1),$$

where the last estimate was done above in connection with the average order of $\sigma(n)/n$. So far, we’ve proved that

$$\sum_{n \leq x} \frac{\varphi(n)}{n} = \beta x + O(\log x), \quad \beta = \sum_{d=1}^{\infty} \frac{\mu(d)}{d^2}.$$

We’d like to figure out the sum we’ve called $\beta$. Note that if

$$\prod_p \left(1 - \frac{1}{p^2}\right)$$

is multiplied out, then each term is of the form $1/d^2$ with $d$ squarefree and the sign of the term depends on whether $d$ has an even or odd number of prime factors. Namely, this infinite product is equal to $\beta$. Changing a sum to a product like this, using unique factorization into primes is an important technique, it’s called an Euler product. Let’s try it out in a slightly more complicated example:

$$\prod_p \left(1 + \frac{1}{p^2} + \frac{1}{p^4} + \frac{1}{p^6} + \ldots\right).$$

When this is multiplied out we see terms of the form $1/n^2$ for every positive integer $n$, once and only once. Thus, this product is exactly

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

which we know after Euler to be $\pi^2/6$. However, note that the infinite sum that appears in each factor of the product just above can be summed as a geometric progression to $1/(1 - 1/p^2)$. That is,

$$\frac{\pi^2}{6} = \sum_{n=1}^{\infty} \frac{1}{n^2} = \prod_p \frac{1}{1 - \frac{1}{p^2}}.$$
and guess what?, this is exactly $1/\beta$. Thus,

$$
\beta = \sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} = \prod_p \left( 1 - \frac{1}{p^2} \right) = \frac{6}{\pi^2}.
$$

We can now complete the computation of the average ordre of $\varphi(n)$ using partial summation. We have

$$
\sum_{n \leq x} \varphi(n) = \sum_{n \leq x} \frac{\varphi(n)}{n} n = x \sum_{n \leq x} \frac{\varphi(n)}{n} - \int_1^x \sum_{n \leq t} \frac{\varphi(n)}{n} dt.
$$

The first term works out to

$$
\frac{6}{\pi^2} x^2 + O(x \log x)
$$

and the second term works out to

$$
-\frac{3}{\pi^2} x^2 + O(x \log x),
$$

so combining, we have that

$$
\frac{1}{x} \sum_{n \leq x} \frac{\varphi(n)}{n} = \frac{3}{\pi^2} x + O(\log x).
$$

The takeaways here: we have introduced two interesting functions, $\varphi$ and $\mu$, we’ve discussed Euler products, and we’ve seen once again that it is desirable to write a function as a sum over divisors of the argument, not to mention the utility of partial summation.

4 Convolution of arithmetic functions

For arithmetic functions $f, g$ we can combine them in some boring ways, such as, $f + g$ and $fg$. This is pointwise addition and pointwise multiplication. Sometimes we can take a composition. But we always can do a convolution, also called Dirichlet multiplication. We define the arithmetic function $f \ast g$ as

$$(f \ast g)(n) = \sum_{d | n} f(d)g(n/d).$$

Using the duality between divisors $d$ of $n$ and $n/d$, we have

$$(f \ast g)(n) = \sum_{d | n} f(n/d)g(d),$$

so that $f \ast g = g \ast f$. A “fraction-free” way to write this:

$$(f \ast g)(n) = \sum_{ab = n} f(a)g(b).$$
One can show that convolution obeys the associative law. In fact,

\[(f * g) * h)(n) = (f * (g * h))(n) = \sum_{abc=n} f(a)g(b)h(c),\]

as is easily verified.

Let \(u\) denote the arithmetic function which is identically 1. And let \(E\) denote the function with \(E(n) = n\) for all \(n\). (Thus, \(u\) is the identity function for pointwise multiplication and \(E\) is the identity function for composition.) We can record some formulas we used above, but now using convolution notation:

\[\tau = u * u, \quad \sigma = E * u, \quad \varphi = \mu * E.\]

Another interesting function is \(I\) defined as

\[I(n) = \left\lfloor \frac{1}{n} \right\rfloor = \begin{cases} 1, & n = 1, \\ 0, & n > 1. \end{cases}\]

It is the identity function for convolution:

\[f * I = f\]

for all arithmetic functions \(f\), as can be easily checked.

Which arithmetic function have inverses under convolution? One example:

\[\mu * u = I.\]

Do you believe this? We’re saying that for any integer \(n > 1\) we have

\[\sum_{d|n} \mu(d) = 0,\]

and for \(n = 1\) the sum is 1. The latter is obvious. To see the former, note that if \(n\) has exactly \(k\) distinct prime divisors, then it has \(\binom{k}{2}\) squarefree divisors of the form \(pq\), it has \(\binom{k}{3}\) squarefree divisors of the form \(pqr\), etc., where the \(p, q, r, \ldots\) are chosen from the prime factors of \(n\). Thus,

\[\sum_{d|n} \mu(d) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} = (1 - 1)^k = 0,\]

when \(k > 0\), using the binomial theorem.

We have the following important corollary:

**Theorem 1** (Möbius inversion formula). If \(f\) is an arithmetic function, then \(g = \mu * f\) satisfies \(u * g = f\). Further, if \(g\) is an arithmetic function satisfying \(u * g = f\), then \(g = \mu * f\).

**Proof.** Say \(g = \mu * f\). We have \(u * g = u * (\mu * f) = (u * \mu) * g = I * f = f\). Conversely, if \(f = u * g\), we have \(\mu * f = \mu * (u * g) = (\mu * u) * g = I * g = g\). □

So if you’d like to write \(f(n)\) in the form \(\sum_{d|n} g(d)\) for some function \(g\), you don’t have far to look, you should take \(g\) as the function \(\mu * f\).
5 Multiplicative functions

Some arithmetic functions $f$ satisfy the identity

$$f(mn) = f(m)f(n) \text{ when } \gcd(m, n) = 1.$$ 

If we also have that $f$ is not the 0-function, we say $f$ is multiplicative. Note that the functions $u, E, \mu$, and $I$ are all easily seen to be multiplicative. (With $u, E, I$ it is not necessary for $m, n$ to be coprime in the identity, they are “completely multiplicative”. However, $\mu$ is not completely multiplicative, just multiplicative.)

A fact about multiplicative functions is that $f(1) = 1$, can you prove it?

We have the following important result.

**Theorem 2.** If $f, g$ are multiplicative functions, so is $f \ast g$.

**Proof.** We assume that $f, g$ are multiplicative and that $\gcd(m, n) = 1$. We have

$$(f \ast g)(mn) = \sum_{ab=mn} f(a)g(b).$$

Since $\gcd(m, n) = 1$, the equation $ab = mn$ implies that if we let

$$a_1 = \gcd(a, m), \ a_2 = \gcd(a, n), \ b_1 = \gcd(b, m), \ b_2 = \gcd(b, n),$$

then $a_1a_2 = a, b_1b_2 = b, a_1b_1 = m, a_2b_2 = n$. Think about it! Thus,

$$(f \ast g)(mn) = \sum_{a_1b_1=m, a_2b_2=n} f(a_1a_2)g(b_1b_2) = \sum_{a_1b_1=m, a_2b_2=n} f(a_1)f(a_2)g(b_1)g(b_2),$$

where we’ve used that $f, g$ are multiplicative and that $\gcd(a_1, a_2) = 1, \gcd(b_1, b_2) = 1$. This last quadruple sum over $a_1, a_2, b_1, b_2$ factors as

$$\sum_{a_1b_1=m} f(a_1)g(b_1) \sum_{a_2b_2=n} f(a_2)g(b_2) = (f \ast g)(m)(f \ast g)(n).$$

Thus, $f \ast g$ is multiplicative. □

It is easy to go from the definition of a multiplicative function to this more general property a multiplicative function $f$ enjoys: If $m_1, m_2, \ldots, m_k$ are positive integers that are pairwise coprime, then

$$f(m_1m_2\ldots m_k) = f(m_1)f(m_2)\ldots f(m_k).$$

In particular, if $n$ has the prime factorization $p_1^{a_1}p_2^{a_2}\ldots p_k^{a_k}$, where $p_1, p_2, \ldots, p_k$ are different primes and $a_1, a_2, \ldots, a_k$ are positive integers, then

$$f(n) = f(p_1^{a_1}) f(p_2^{a_2}) \ldots f(p_k^{a_k}).$$
This is telling us that if you know that $f$ is multiplicative and if you know what $f$ does to primes and powers of primes, then you know what $f$ does to every number. This principle has some cool applications. For example, we saw above that $\mu * u = I$, and we proved this by a combinatorial argument and the binomial theorem. Consider now this proof: We know that $\mu$ and $u$ are multiplicative functions, as is $I$. To see that $\mu * u$ is equal to $I$ it suffices to show that they do the same thing to prime powers. If $p$ is a prime and $a$ is a positive integer, we have (since $\mu$ is 0 on non-squarefrees)

$$(\mu * u)(p^a) = \sum_{d | p^a} \mu(d) = \mu(1) + \mu(p) = 1 - 1 = 0,$$

and of course $I(p^a) = 0$, so they are equal functions.

Here is another illustration. Find what $\varphi * u$ is equal to. Call this convolution $f$. We know that $f$ is multiplicative, since both $\varphi$ and $u$ are. Lets compute $f(p^a)$. We have

$$f(p^a) = \sum_{d | p^a} \varphi(d) = \sum_{i=0}^{a} \varphi(p^i) = 1 + (p - 1) + (p^2 - p) + \cdots + (p^a - p^{a-1}) = p^a,$$

since this sum “telescopes”. Thus, $f$ and $E$ agree at prime powers, so we have $f = E$. That is, $\varphi * u = E$:

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

We can see this relation two other ways. First, since we proved that $\varphi = \mu * E = E * \mu$, it follows that $\varphi * u = E * \mu * u = E * I = E$. The second proof uses no fancy notation. Take the $n$ fractions $1/n, 2/n, \ldots, n/n$ and reduce them to lowest terms. In lowest terms, a denominator will be a divisor $d$ of $n$, and every fraction $a/d$ with $1 \leq a \leq d$ and gcd$(a, d) = 1$ will appear, and no others. Hence there are $\varphi(d)$ of the original fractions that reduce to denominator $d$. This means that if we count the total number of reduced fractions, which is the total number of fractions we start with, which is $n$, we get that it is also equal to

$$\sum_{d | n} \varphi(d).$$