AVERAGE VALUES CONTINUED: φ AND μ

1. AVERAGE ORDER OF THE TOTIENT FUNCTION, AND APPLICATIONS

Theorem 1. *The average order of the totient function is $g(n) = \frac{1}{2\zeta(2)}n = \frac{3}{\pi^2}n$.*

It is nice to have a "harder analysis" analogue of Theorem 1. That is, the theorem at the moment asserts that $\frac{1}{n}\sum_{k=1}^{n}\varphi(k) \sim \frac{3}{\pi^2}n$, or equivalently

$$\lim_{n\to\infty} \frac{\sum_{k=1}^{n}\varphi(k)}{\frac{3}{\pi^2}n^2} = 1.$$

This in turn means that if we define "the error term" $E(n) = \sum_{k=1}^{n}\varphi(k) - (\frac{3}{\pi^2})n^2$, so that $\sum_{k=1}^{n}\varphi(k) = \frac{3}{\pi^2}n^2 + E(n)$, then the error term is small compared to the main term: namely it is equivalent to

$$\lim_{n\to\infty} \frac{E(n)}{(3/\pi^2)n^2} = 0.$$

So far we are just pushing around the definitions. But a fundamentally better thing to do would be to give an upper bound on the error $E(n)$, i.e., to find a nice, simple function $h(n)$ such that $E(n) \le h(n)$ and $\frac{h(n)}{(3/\pi^2)n^2} \to 0$ In fact we do not need $E(n) \le h(n)$ quite: if $E(n)$ were less than $100h(n)$ that would be just as good, because if $h(n)$ divided by $(3/\pi^2)n^2$ approaches zero, the same would hold for $100h(n)$. This motivates the following notation:

We say that $f(n) = O(g(n))$ if there exists a constant C such that for all n, $f(n) \le Cg(n)$. So it would be enough to show that $E(n) = O(h(n))$ for some function h which approaches zero when divided by $\frac{3}{\pi^2}n^2$, or in more colloquial language, for any function grows less than quadratically. So for instance a stronger statement would be that $E(n) = O(n^\delta)$ for some $\delta < 2$. In fact one can do a bit better than this:

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

Or again, dividing through by n, this theorem asserts that that the average over the first n values of the ϕ function is very nearly $\frac{3}{\pi^2}n$, and the precise sense in which this is true is that the difference between the two is bounded by a constant times $\log n$. Note that it would be best of all to know an actual acceptable value of such a constant – that would be a "completely effective" version of the statement, but we do not go into this here (or need to).

Now we can give a very nice application, which we can first state in geometric language. It concerns the lattice points in the plane. Suppose we are some lord of the lattice points, sitting at the origin $(0,0)$ and surveying our entire domain. What do we see? Well – assuming a truly 2-dimensional situation – we can see

some of our loyal subjects but not others. For instance we can certainly see the point $(1,1)$, but we cannot see $(2,2)$, $(3,3)$ or any $n(1,1)$ with $n > 1$ since the point $(1,1)$ itself obscures our view.

Thus we can define a lattice point (x,y) to be **visible** (from the origin) if the line segment from $(0,0)$ to (x,y) contains no lattice points on its interior. Suppose we start coloring the lattice, coloring a lattice point red if we can see it and black if we cannot see it. Try it yourself: this gives a very interesting pattern. It is natural to ask: how many of the lattice points can we see?

Well, the first observation is that a lattice point (x,y) is visible iff $\gcd(x,y) = 1$: an obstructed view comes precisely from a nontrivial common divisor of x and y. From this it follows that the answer is "infinitely many": for instance we can see $(1,n)$ for all integers n, and many more besides. Well, let us change our question a bit. Suppose that each of the lattice points is supposed to pay a flat tax to our lordship, and if we see a lattice point then we can see whether or not it has paid its taxes. What percentage of our revenue are we collecting if we only worry about the lattice points we can see?

Now to formalize the question. If we ask about the entire lattice at once, the answer to most of our questions is always going to be "infinity," and moreover an actual king (even a two-dimensional one) probably rules over a finite kingdom. So for a positive integer N, let us write $L(N)$ for the number of lattice points (x,y) with $|x|$, $|y| \leq N$ – that is, the lattice points lying in the square centered at the origin with length (and width) equal to $2N$. Well, this number is $(2N+1)^2$: there are $2N+1$ possible values for both x and y. But now define $V(N)$ to be the number of visible lattice points, and our question is: when N is large, what can we say about $\frac{V(N)}{L(N)}$?

Theorem 3. *We have* $\lim_{N\to\infty} \frac{V(N)}{L(N)} = \frac{6}{\pi^2}$.

Before we prove the result, we can state it in a slightly different but equally striking way. We are asking after all for the number of ordered pairs of integers (x,y) each of absolute value at most N, with x and y relatively prime. So, with a bit of poetic license perhaps, we are asking: what is the probability that two randomly chosen integers are relatively prime? If we lay down the ground rules that we are randomly choosing x and y among all integers of size at most N, then the astonishing answer is that we can make the probability as close to $\frac{6}{\pi^2}$ as we wish by taking N sufficiently large.

Now let us prove the result, or at any rate deduce it from Theorem 1. First we observe that the eight lattice points immediately nearest the origin – i.e., those with $\max(|x|,|y|) \leq 1$ – are all visible. Indeed there is an eightfold symmetry in the situation: the total number of visible lattice points in the square $|x|, |y| \leq N$ will then be these 8 plus 8 times the number of lattice points with $2 \leq x \leq N$, $1 \leq y \leq x$ (i.e., the ones whose angular coordinate θ satisfy $0 < \theta \leq \frac{\pi}{2}$. But now we have
$$V(N) = 8 + \sum_{2 \leq n \leq N} \sum_{1 \leq m \leq n, (m,n)=1} 1 = 8 \sum_{1 \leq n \leq N} \varphi(N).$$

But now $L(N) = (2N+1) \sim 4N^2$, and $8\sum_{n=1}^{N} \varphi(n) \sim \frac{3}{\pi^2} N^2$, so that

$$\frac{V(N)}{L(N)} = \frac{8}{(2N+1)^2} + 8\frac{\sum_{n=1}^{N} \varphi(n)}{(2N+1)^2} \to 0 + 8 \cdot (\frac{3}{\pi^2})/4 = \frac{6}{\pi^2}.$$

Remark: In the proof we used the following fact on asymptotic functions: if $P_1(N) \sim P_2(N)$ and $Q_1(N) \sim Q_2(N)$, then $\lim_{N \to \infty} \frac{P_1(N)}{Q_1(N)} = \lim_{N \to \infty} \frac{P_2(N)}{Q_2(N)}$. You should stop and check this (it's not hard) if you've never seen it before.

Alternate "proof": the following argument is not in itself rigorous, but is faster and more interesting.[1] Namely, what does it "really mean" for two integers x and y to be relatively prime? It means that there is no prime number p which simultaneously divides both x and y. Remarkably, this observation leads directly to the result. Namely, the "chance" that x is divisible by a prime p is evidently $\frac{1}{p}$, so the chance that x and y are both divisible by p (independent events!) is $\frac{1}{p^2}$. Therefore the chance that x and y are *not* both divisible by a prime p is $(1 - p^{-2})$. (In fact we have not yet used the fact that p is prime.) Now we think of being divisible by different primes as again being independent events: if I tell you that an integer is divisible by 3, not divisible by 5 and divisible by 7, and ask you what are the odds that it is divisible by 11, then we still think the chance is $\frac{1}{11}$. Now the probability that each of a set of independent events all occur is the product of the probabilities that each of them occur, so the probability that x and y are not simultaneously divisible by any prime p ought to be $(1-2^{-2}) \cdot (1-3^{-2}) \cdot \ldots (1-p^{-2}) \cdots$, but we saw earlier that this infinite product is nothing else than the reciprocal of $\sum_{n=1}^{\infty} \frac{1}{n^2} = \zeta(2)$. Thus the answer should be $\frac{1}{\zeta(2)} = \frac{6}{\pi^2}$!

The argument was not rigorous because the integers are not really a probability space: there is nothing random about whether, say, 4509814091046 is divisible by 103; either it is or it isn't. Instead of probability one should rather work with the notion of the "density" of a set of integers (or of a set of pairs of integers) – a notion which we shall introduce rather soon – and then all is well until we pass from sets defined by divisibility conditions on finitely many primes to divisibility conditions on all (infinitely many) primes. This is not to say that such a "probability-inspired proof" cannot be pushed through – it absolutely can. Moreover, the fact that the probabilistic argument gives an answer which can be proven to be the correct answer via conventional means is perhaps most interesting of all.

Finally, we note that probabilistic reasoning gives the same answer to a closely related question: what is the probability that a large positive integer n is squarefree? This time we want, for each prime p "independently", n not to be divisible by p^2, of which $1 - p^{-2}$ percent of all integers are. Therefore we predict that the probability that n is squarefree is also $\frac{6}{\pi^2}$, and this too can be proved by similar (although not identical) means to the proof of Theorem 1.

---

[1] Of course it can be *made* rigorous, or I wouldn't waste your time with it.

2. The average order of the Möbius function

We are interested in the behavior of

$$\mu_a(n) = \frac{1}{n} \sum_{k=1}^{n} \mu(k).$$

This is a horse of a completely different color, as we are summing up the values 0 and ± 1. We just saw that μ is nonzero a positive proportion, namely $\frac{6}{\pi^2}$, of the time. Looking at values of the Möbius function on squarefree integers one finds that it is indeed $+1$ about as often as it is -1, which means that there ought to be a lot of cancellation in the sum. If every single term in the sum were 1 then $\mu_a(n)$ would still only be equal to 1, and similarly if every single term were -1 the average order would be -1, so the answer (if the limit exists!) is clearly somewhere in between.

Let's think a little more. Restricting to squarefree numbers (as a $\frac{6}{\pi^2}$ proportion are), it is not hard to believe that relatively few integers are divisible only by a fixed number of primes (we will prove a few results in this direction later on): in other words, for large N, most squarefree integers $1 \leq n \leq N$ will have lots of prime factors, and guessing whether they have an even or odd number of factors seems like guessing whether a large number is even or odd: without any further information, the most obvious guess is that $\mu(n) = +1$ about as often as it equals -1. The following theorem makes this quantitative:

Theorem 4. *The average order of the Möbius function is zero:*

$$\lim_{n \to \infty} \frac{\sum_{k=1}^{n} \mu(k)}{n} = 0.$$

However, this turns out to be a formidably deep result. Namely, it has been known for almost two hundred years that this statement is logically equivalent to the single greatest result in analytic number theory: the prime number theorem (coming up!). However the prime number theorem has only been known to be true for a bit over one hundred years.

In fact one can ask for more: the statement that the average order of μ is zero is equivalent to the statement that the sum $\sum_{k=1}^{n} \mu(k)$ is of smaller order than n itself, i.e., we need just enough cancellation to beat the trivial bound. But if you do some computations you will see that these partial sums seem in practice to be *quite a bit* smaller than n, and to say exactly how large they should be turns out to be a much deeper problem yet: it is equivalent to the **Riemann hypothesis**. I hope to return to this later, but for now I leave you with the question: suppose you believed that the nonzero values of the Möbius function were *truly* random: i.e., for every n we flip a coin and bet on heads: if we win, we add 1 to the sum, and if we lose, we subtract 1 from the sum. It is then clearly ridiculous to expect to win or lose all the games or anything close to this: after n games we should indeed be much closer to even than to having won or lost n dollars. But how close to even should we expect to be?