Math 229: Introduction to Analytic Number Theory

Elementary approaches I: Variations on a theme of Euclid

Like much of mathematics, the history of the distribution of primes begins with Euclid:

Theorem (Euclid [IX, 20])**.** *There are infinitely many primes.*

Euclid's justly famed argument, while often presented as a proof by contradiction, is readily framed as an effective (albeit rather inefficient) construction:

Proof: We inductively construct a sequence $\{p_k\}_{k>1}$ of primes, starting with $p_1 = 2$. Given primes $p_1, p_2, \ldots, p_n$, let $P_n = \prod_{k=1}^n p_k$, define $N_n = P_n + 1$, and let p_{n+1} be the smallest factor of N_n. Then p_{n+1} is a prime no larger than N_n and different from $p_1, \ldots, p_n$. Thus $\{p_k\}_{k>1}$ is an infinite sequence of distinct primes, Q.E.D.

This answers Yes to the first asymptotic question to ask about

$$\pi(x) := \#\{p \leq x : p \text{ is a positive prime}\} = \sum_{\substack{0<p\leq x \\ p \text{ prime}}} 1,$$

namely whether $\pi(x) \to \infty$ as $x \to \infty$. Moreover, the proof also gives an explicit upper bound on p_n, and thus a lower bound on $\pi(x)$.

Theorem. *For each integer $n > 0$, there are more than n primes $p < 2^{2^n}$. Equivalently, we have*[1]

$$\pi(x) > \log_2 \log_2 x$$

for all $x > 1$.

Proof: In the proof of Euclid's theorem, we may take $p_1 = 2$, and observe that

$$p_{n+1} \leq N_n = 1 + \prod_{k=1}^n p_k \leq 2 \prod_{k=1}^n p_k.$$

if equality were satisfied at each step we would have $p_n = 2^{2^{n-1}}$. Thus by induction we see that

$$p_n \leq 2^{2^{n-1}},$$

and of course the inequality is strict once $n > 1$. Therefore if $x \geq 2^{2^{n-1}}$ then $p_k < x$ for $k = 1, 2, \ldots, n$, and so $\pi(x) \geq n$, Q.E.D.

The $P_n + 1$ trick has been adapted to prove some special cases of Dirichlet's theorem on primes in arithmetic progressions, which asserts that for relatively

[1] The ubiquity of estimates involving $\log\log x$, $\log\log\log x$, and beyond in analytic number theory has even inspired a joke:

> Q: What sound does a drowning analytic number theorist make?
> A: log log log log . . .

I do not know who originated this joke; I was told it by B. Mazur, who heard it from R. Murty.

prime integers $q > 0$ and a there are infinitely many primes $p \equiv a \bmod q$. (We shall give the proof later in the course.) Of course the case 1 mod 2 is trivial given Euclid. For $-1 \bmod q$ with $q = 3, 4, 6$, start with $p_1 = q - 1$ and define $N_n = qP_n - 1$. More generally, for any quadratic character[2] χ there are infinitely many primes p with $\chi(p) = -1$; as a special case, given an odd prime q_0, there are infinitely many primes p which are quadratic nonresidues of q_0. [I'm particularly fond of this argument because I was able to adapt it as the punchline of my doctoral thesis; see [Elkies 1987].] The case of $\chi(p) = +1$ is only a bit trickier.[3] For instance, to prove Dirichlet for $(q, a) = (4, 1)$, let $p_1 = 5$ and $N_n = 4P_n^2 + 1$, and invoke Fermat's theorem on the prime factors of $x^2 + y^2$. Again this argument even yields an explicit lower bound on

$$\pi(x, 1 \bmod 4) := \#\{p \leq x : p \text{ is a positive prime congruent to 1 mod 4}\},$$

namely[4]

$$\pi(x, 1 \bmod 4) > C \log\log x$$

for some positive constant C.

But Euclid's approach and its variations, however elegant, are not sufficient for our purposes. For one thing, numerical evidence suggests — and we shall soon prove — that $\log_2 \log_2 x$ is a gross underestimate on $\pi(x)$. For another, one cannot prove all cases of Dirichlet's theorem using only variations on the Euclid argument.[5] The elementary approaches to lower bounds on $\pi(n)$ in the next few lectures will address at least the first deficiency.

Meanwhile we briefly consider upper bounds. Trivially $\pi(x) < x$ for all x. Since all primes except 2 are odd (and 1 is not prime), we have in fact $\pi(x) \leq (x+1)/2$, and $\pi(2m) \leq m$ for $m = 1, 2, 3, \ldots$. Likewise since all but two primes are congruent to 1 or 5 mod 6 we have $\pi(6m) \leq 2m + 1$ for natural numbers m, and so $\pi(x) \leq (x/3) + c_2$ for some constant c_2. More generally, let p_n be the n-th prime, and let[6] $P_n = \prod_{k=1}^{n} p_k$; then by the Chinese Remainder Theorem only $\prod_{k=1}^{n}(p_k - 1)$ of the residue classes mod P_n can contain any prime other than $p_1, \ldots, p_n$, whence

$$\pi(P_n m) < \left(\prod_{k=1}^{n}(p_k - 1)\right) m + n$$

[2] We'll say much more about characters, quadratic and others, in a few weeks. An example is the Legendre symbol $n \mapsto (n/q_0)$ for some fixed odd prime q_0, or more generally any nonempty product $n \mapsto \prod_{i=1}^{s}(n/q_i)$ of such characters for distinct odd primes $q_1, \ldots, q_s$.

[3] But enough so that a problem from a recent Qualifying Exam for our graduate students asked to prove that there are infinitely many primes congruent to 1 mod 4.

[4] Even a drowning analytic number theorist knows that $\log\log$ and $\log_2 \log_2$ are asymptotically within a constant factor of each other. What is that factor?

[5] This is <u>not</u> a theorem, of course. How could one even define "variation of the Euclid argument" rigorously? But a Euclid-style argument for the infinitude of primes congruent to 2 mod 5 or mod 7 would already be quite impressive.

[6] By analogy with the "factorial" $n! = \prod_{i=1}^{n} i$, this P_n is sometimes called the n-th "primorial".

for all natural numbers m. Thus there is a constant c_n such that

$$\pi(x) < \left(\prod_{k=1}^{n} \frac{p_k - 1}{p_k} \right) x + c_n \tag{1}$$

for all x. What happens to this upper bound as $n \rightarrow \infty$? The coefficient of x decreases, and if it approaches zero — that is, if $\prod_{k=1}^{\infty} \bigl((p_k - 1)/p_k\bigr) = 0$ — then $\pi(x)/x \rightarrow 0$ as $x \rightarrow \infty$. Otherwise we might expect that $\pi(x)/x$ approaches some positive constant, say κ. But then $\prod_{k=1}^{\infty} \bigl((p_k - 1)/p_k\bigr)$ should certainly vanish, because it contains a positive proportion of the factors of the infinite product $\prod_{n=2}^{\infty}((n-1)/n)$, whose partial products are

$$\prod_{n=2}^{N} \frac{n-1}{n} = \frac{(N-1)!}{N!} = \frac{1}{N} \rightarrow 0.$$

Unfortunately it does not immediately follow from (1) that if $\prod_{k=1}^{\infty} \bigl((p_k - 1)/p_k\bigr)$ is positive then it equals $\lim_{x\rightarrow\infty} \pi(x)/x$; indeed for all we know so far the limit might not exist at all, and we can conclude only that

$$\limsup_{x\rightarrow\infty} \frac{\pi(x)}{x} \leq \prod_{k=1}^{\infty} \frac{p_k - 1}{p_k}.$$

(See the next-to-last Exercise below!) But we can still use (1) to obtain an elementary proof that $\pi(x)/x \rightarrow 0$ as $x \rightarrow \infty$. Assume not. Then there exists $\kappa > 0$ and an infinite sequence $x_m \rightarrow \infty$ such that $\pi(x_m) > \kappa x_m$ for all $m \geq 1$. Passing to a subsequence we may assume that the x_m grow so rapidly that $x_{m-1} < (\kappa/2) x_m$ for all $m > 1$. Then more than $(\kappa/2) x_m$ of the primes $p \leq x_m$ are in $(x_{m-1}, x_m]$. Hence the product of $(p-1)/p$ over these primes is $< 1 - (\kappa/2)$ (why?), whence $\prod_{p \leq x_m}((p-1)/p) < (1 - (\kappa/2))^m$. But then $\prod_{k=1}^{\infty} \bigl((p_k - 1)/p_k\bigr) = 0$, and we already know that this implies the desired conclusion $\pi(x)/x \rightarrow 0$.

Notice that we have not yet proved that in fact $\prod_{k=1}^{\infty} \bigl((p_k - 1)/p_k\bigr) = 0$. That is in fact true, and will be part of our next topic.

Exercises

1. Fix an integer $q > 2$, and let G be a subgroup of $(\mathbf{Z}/q\mathbf{Z})^*$ other than $(\mathbf{Z}/q\mathbf{Z})^*$ itself. Prove that there are infinitely many primes whose residue modulo q is not in G.

2. Obtain explicit values of C and x_0 such that $\pi(x, 1 \bmod 4) > C \log\log x$ for all $x > x_0$.

3. Use cyclotomic polynomials to show more generally that for any q_0, prime or not, there exist infinitely many primes congruent to 1 mod q_0. [Attributed to Euler in [Dickson 1919, Ch.XVIII], a chapter which gives much more information on the history of work on the distribution of primes up to about 1900. Note

that $4P_n^2+1$ is the fourth cyclotomic polynomial evaluated at $2P_n$.] Show that again the number of such primes $\leq x$ grows at least as fast as some multiple of $\log\log x$ (that is, for every q_0 there exist constants $C>0$ and x_0 such that for every $x>x_0$ there are at least $C\log\log x$ primes $p\leq x$ congruent to 1 mod q_0).

4. Show that there are infinitely many primes congruent to 4 mod 5. If you know about the arithmetic of cyclotomic number fields, prove more generally that if G_0 is a subgroup of $(\mathbf{Z}/q\mathbf{Z})^*$ and G is a subgroup of G_0 other than G_0 itself then there are infinitely many primes whose residue mod q is in G_0 but not in G. Obtain an explicit $\log\log$ bound as before.

5. [A much later proof of the infinitude of primes that curiously gives the same bound $\pi(x)>\log_2\log_2(x)$.] Recall that the m-th Fermat number F_m is defined by $F_m=2^{2^m}+1$ $(m=0,1,2,\ldots)$. Prove that F_m and $F_{m'}$ are relatively prime unless $m=m'$. Conclude that there are at least n primes $p\leq F_{n-1}$, and thus that $\pi(x)>\log_2\log_2 x$.

6. [A warning about limits.] Let $\{q_k\}_{k=1}^{\infty}$ be a sequence of primes, or even of natural numbers that are relatively prime in pairs; set $\kappa=\prod_{k=1}^{\infty}\bigl((q_k-1)/q_k\bigr)$; and let $\{a_k\}_{k=1}^{\infty}$ be an arbitrary sequence of integers. For $x>0$ let $\varpi(x)$ be the number of positive integers $q\leq x$ such that $q\not\equiv a_k \bmod q_k$ for all k. Prove that $\limsup_{x\to\infty}\varpi(x)/x\leq\kappa$, but that there are choices of $\{a_k\}$ for which $\varpi(x)/x\to 0$ as $x\to\infty$, even if $\kappa>0$.

7. [Analytic number theory without Euclid.] Even though we're starting our exposition with Euclid, there are questions and techniques that can precede even the infinitude of primes. An example is lower bounds on Euler's phi-function

$$\phi(q)=\#\bigl((\mathbf{Z}/q\mathbf{Z})^*\bigr)=\#\{a\mid 0\leq a<q,\ \gcd(a,q)=1\}=q\prod_{p|q}\frac{p-1}{p}$$

(q a positive integer). Clearly $\phi(q)\leq q$, with equality only for $q=1$; but often (in analytic number theory and elsewhere) it is useful to know that $\phi(q)$ is not too much smaller than q, either for any given q or on average.
i) Prove that for all $\epsilon>0$ there exists $C_\epsilon>0$ such that $\phi(q)>C_\epsilon q^{1-\epsilon}$ for all q. [Only finitely many of the factors $(p-1)/p$ can be smaller than $p^{-\epsilon}$.]
ii) Prove that there exists $c>0$ such that $\sum_{q=1}^{x}\phi(x)>cx^2$ for all positive integers x. [Since $\sum_{q=1}^{x}q=(x^2+x)/2$, the lower bound on $\sum_{q=1}^{x}\phi(x)$ says in effect that $\phi(q)\geq 2cq$ on average.]
Naturally, more precise results on the distribution of primes will let us obtain sharper results, such as the dependence of C_ϵ on ϵ; e.g. the fact that C_ϵ cannot be chosen independent of ϵ is tantamount to $\liminf_q(\phi(q)/q)=0$, which is in turn equivalent to $\prod_{k=1}^{\infty}\bigl((p_k-1)/p_k\bigr)=0$. For part (ii), it is easy to obtain this bound with $2c=1-\sum_{k=1}^{\infty}(1/p^2)$. The actual asymptotic average of $\phi(q)/q$ is usually described as the probability that two randomly chosen integers are relatively prime. This is well-known to be $1/\zeta(2)=6/\pi^2$; this is not too hard to prove (see Exercise 7 of the next installment), though there are still nontrivial questions concerning the error term $\sum_{q=1}^{x}\phi(x)-x^2/(2\zeta(2))$.

Digression

Even a piece of mathematics as venerable as Euclid's proof of the infinitude of primes can continue to suggest very difficult problems. For instance, let p_n be the n-th prime and $P_n = \prod_{k=1}^{n} p_k$ as before. We know that $P_n + 1$ must contain a new prime factor, which cannot be p_{n+1} once $n > 1$ (if only because $P_n - 1$ must also contain a new prime factor). Does it happen infinitely often that p_{n+1} is a factor of $P_n + 1$? [This is the case for $n = 1$, 7, 232, 430, and no other $n < 10^5$.] What of the primality of $P_n + 1$ itself? It is well-known that $P_n + 1$ is prime for $n = 1, 2, 3, 4, 5$, but $P_6 + 1 = 30031 = 59 \cdot 509$. As of early 2010, only seventeen $n > 5$ have been found for which $P_n + 1$ is prime, of which the smallest is 11 and the largest is 33237.[7] Again it is not known whether this happens infinitely often. Likewise for the primality of $P_n - 1$ and its divisibility by p_{n+1}. For another variation, define $q_1 = 2$ and, for $n > 0$, let q_{n+1} be the smallest prime factor of $(\prod_{k=1}^{n} q_k) + 1$. The sequence $\{q_n\}_{n=1}^{\infty}$ starts

$$2, 3, 7, 43, 13, 53, 5, 6221671, 38709183810571, 139, 2801, 11, \ldots$$

For instance, $q_5 = 13$ because $2 \cdot 3 \cdot 7 \cdot 43 + 1 = 1807 = 13 \cdot 139$. Is this "Euclid-Mullin sequence" [Sloane, A000945] a permutation of the sequence of primes? Probably yes, but proving this will likely be intractable for the foreseeable future. The same is true for the infinitude of primes of the form $P_n \pm 1$, and of n such that $p_{n+1} | P_n \pm 1$.

It should not even be obvious that one should expect that these four sets are all infinite. The heuristics supporting this expectation rely on results on the distribution of primes that we shall develop in the next few weeks.

References

[Dickson 1919] Dickson, L.E.: *History of the Theory of Numbers, Vol. I: Divisibility and Primality.* Washington: Carnegie Inst., 1919.

[Euclid] Euclid, *Elements.*

[Elkies 1987] Elkies, N.D.: The existence of infinitely many supersingular primes for every elliptic curve over **Q**, *Invent. Math.* **89** (1987), 561–568; See also: Supersingular primes for elliptic curves over real number fields, *Compositio Math.* **72** (1989), 165–172.

[Sloane] Sloane, N.J.A.: *On-Line Encyclopedia of Integer Sequences.*
`http://www.research.att.com/~njas/sequences`

[7] Sequence A014545 in [Sloane], where the primality of $P_{13494} + 1$ is attributed to Eric W. Weisstein, March 13, 2004. For the analogous question concerning $P_n - 1$, see Sequence A055704 and A006794.