1 Primes

Primes would seem to be the ultimate in precision. A number 317 is either prime or it isn't (this one is!), there is no approximation to its primality. Nonetheless, Asymptopia is the proper place to examine primes in the aggregate.

Definition 1 *For* $n \geq 2$, $\pi(n)$ *denotes the number of primes* p *with* $2 \leq p \leq n$.

Our goal in this chapter is to show one of the great theorems of mathematics.

Theorem 1.1 (The Prime Number Theorem)

$$\pi(n) \sim \frac{n}{\ln n} \tag{1}$$

This result was first conjectured in the early nineteenth century. (While the conjecture is sometimes attributed to Gauss the history is murky.) It was a central problem for that century, finally being proven independently by Hadamard and Vallée-Poussin in 1898. There proofs involved complex variables and a long search continued for an elementary proof. This was finally obtained in 1949 by Selberg and Erdős. Still, a full proof of Theorem 1 is beyond the limits of this work. We shall come close to it with the following results:

Theorem 1.2 *There exists a positive constant* c_1 *such that*

$$(c_1 + o(1))\frac{n}{\ln n} \leq \pi(n) \tag{2}$$

That is, $\pi(n) = \Omega(n/\ln n)$. *Further, our argument gives* $c_1 = \ln 2$.

Theorem 1.3 *There exists a positive constant* c_2 *such that*

$$\pi(n) \leq (c_2 + o(1))\frac{n}{\ln n} \tag{3}$$

That is, $\pi(n) = O(n/\ln n)$. *Further, our argument gives* $c_2 = 2\ln 2$.

Together, Theorem 1.2, 1.3 yield:

$$\pi(n) = \Theta(\frac{n}{\ln n}) \tag{4}$$

With more effort we shall show

Theorem 1.4 If *there exists a positive constant c such that*

$$\pi(n) \sim c\frac{n}{\ln n} \tag{5}$$

then $c = 1$.

1.1 Fun with Primes

A Break! No asymptotics in this section!

How many factors of the prime 7 are there in 100!? The numbers $7, 14, \ldots, 98$ all have a factor of 7 so that gives $\frac{98}{7} = 14$ factors. *And,* 49 and 98 have a second factor of 7 which gives an additional $\frac{98}{49} = 2$ factors. In total there are $16 = 14 + 2$ factors of 7.

Definition 2 *For $n \geq 1$ and p prime, $v_p(n)$ denotes the number of factors p in n. Equivalently, $v_p(n)$ is that nonnegative integer a such that p^a divides n but p^{a+1} does not divide n.*

Theorem 1.5 *For any $n \geq 1$ and p prime*

$$v_p(n!) = \sum_{i=1}^{\infty} \lfloor \frac{n}{p^i} \rfloor \tag{6}$$

Equivalently

$$v_p(n!) = \sum_{i=1}^{s} \lfloor \frac{n}{p^i} \rfloor \text{ with } s = \lfloor \log_p n \rfloor \tag{7}$$

When $i > \lfloor \log_p n \rfloor$, $p < n^i$ so the addend in (6), explaining the equivalence. The argument with $p = 7, n = 100$ easily generalizes. For any $i \leq s$ there are $\lfloor np^{-i} \rfloor$ numbers $1 \leq j \leq n$ that have (at least) i factors of p. We count each such i and j once, as then an i with precisely u factors of p will be counted precisely u times.

We apply Theorem 1.5 to study binomial coefficients. Let $n = a + b$ and set $C = \binom{n}{a} = \frac{n!}{a!b!}$. Applying (7)

$$v_p(C) = v_p(n!) - v_p(a!) - v_p(b!) = \sum_{i=1}^{s} \lfloor \frac{n}{p^i} \rfloor - \lfloor \frac{a}{p^i} \rfloor - \lfloor \frac{b}{p^i} \rfloor \tag{8}$$

with $s = \lfloor \log_p n \rfloor$ as in (7).

Theorem 1.6 *With $n = a + b$, p prime, and $C = \binom{n}{a}$,*

$$0 \leq v_p(C) \leq \lfloor \log_p n \rfloor \tag{9}$$

`Proof:` Set $\alpha = ap^{-i}$, $\beta = bp^{-i}$. Then the addend in (8) is

$$\lfloor \alpha + \beta \rfloor - \lfloor \alpha \rfloor - \lfloor \beta \rfloor \tag{10}$$

This term is zero if the fractional parts of α, β sum to less than one and one if they sum to one or more. The sum (8) consists of $s = \lfloor \log_p n \rfloor$ terms, each one or zero, and so lies between 0 and s.

`Remark:` With $n = a + b$ there are two arguments why $a!b!$ divides $n!$. One: the proof of Theorem 8 gives that, for all primes p, $v_p(n!) \geq v_p(a!) + v_p(b!) = v_p(a!b!)$ and thus $a!b!$ divides $n!$. Two: The quotient $\frac{n!}{a!b!} = \binom{n}{a}$ *counts* the a-subsets of an n-sets and hence must be a nonnegative integer. Which proof one prefers is an esthetic question [1] but it is frequently useful to know more than one proof of a theorem.

There is an amusing way of calculating $v_p(C)$ with $C = \binom{n}{a}$ and $a+b = n$. Write a, b is base p. Add them (in base p) so that you will get n in base p.

Theorem 1.7 *$v_p(C)$ is the number of carries when you add a, b getting n, all in base p.*

For example, let $a = 33$, $b = 25$ so $n = 58$ (written in decimal), and set $p = 7$. In base 7, $a = 45$, $b = 34$. When we add them [2]

```
   45
+  34
 ----
  112
```

There we two carries and $v_p(\binom{45}{34}) = 2$.

We indicate the argument. For each $1 \leq i$ we get a carry from the $i-1$-st place (counting from the right, starting at 0) to the i-th place if and only if the fractional parts of ap^{-i} and bp^{-i} add to at least one and that occurs if and only if term (10) is one.

1.2 PMT - Lpper Bound

Let n be even (n odd will be similar). The upper *and* lower bounds come from examining the prime factorization of binomial coefficients. Set $r = \pi(n)$

[1] This author prefers the "counts" argument.

[2] To paraphrase the wonderful songwriter Tom Lehrer, base seven is just like base ten – if you are missing three fingers!

and let $p_1, \ldots, p_r$ denote the primes up to n and write

$$\binom{n}{n/2} = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r} \tag{11}$$

(There might not be a factor of p_i. In that case we simply write $\alpha_i = 0$.) We rewrite the upper bound of Theorem 1.6 as:

$$p_i^{\alpha_i} \leq n \tag{12}$$

Thus

$$\binom{n}{n/2} \leq n^r \tag{13}$$

Stirling's Formula gives an asymptotic formula for $\binom{n}{n/2}$ but here we use only the weaker $\binom{n}{n/2} = 2^{n(1+o(1))}$. Taking ln of both sides of (13) and dividing gives

$$\pi(n) = r \geq \frac{\ln \binom{n}{n/2}}{\ln n} = \frac{n}{\ln n}(\ln 2)(1 + o(1)) \tag{14}$$

What if n is odd? In Asymptopia we simply apply (14) to the even $n - 1$. Thus

$$\pi(n) \geq \pi(n-1) \geq \frac{\ln \binom{n-1}{(n-1)/2}}{\ln(n-1)} \tag{15}$$

which is again $\frac{n}{\ln n}(\ln 2)(1 + o(1))$.

1.3 PMT-Upper Bound

Again assume n is even. There are $\pi(n) - \pi(n/2)$ primes p with $\frac{n}{2} < p < n$. Each of them appears in $\binom{n}{n/2}$ to the first power. (They appear once in the numerator as a factor of p and never in the denominator.) Thus, with the product over these primes,

$$\prod p \leq \binom{n}{n/2} \tag{16}$$

We again do not need a more precise estimate and here simply bound $\binom{n}{n/2} \leq 2^n$. Each factor p is a factor of at least $\frac{n}{2}$. Thus

$$(\frac{n}{2})^{\pi(n)-\pi(\frac{n}{2})} \leq 2^n \tag{17}$$

Taking ln of both sides gives

$$\pi(n) - \pi(\frac{n}{2}) \leq \frac{n}{\ln(n/2)}(\ln 2) \tag{18}$$

For $n = 2k + 1$ odd we apply the same argument to $\binom{n}{k}$ getting an upper bound on $\pi(n) - \pi(k + 1)$. We combine the even and odd cases by writing

$$\pi(n) - \pi(\lceil\frac{n}{2}\rceil) \leq \frac{n}{\ln(n/2)}(\ln 2) \tag{19}$$

Turning (19) into an upper bound on $\pi(n)$ is a typical problem in Asymptopia. Set $x_0 = n$ and $x_{i+1} = \lceil\frac{x_i}{2}\rceil$. This sequence decreases until finally reaching $x_s = 1$. Applying (19) to $n = x_0, \ldots, x_{s-1}$ and adding we get

$$\pi(n) \leq \sum_{i=0}^{s-1} \frac{x_i}{\ln(x_i/2)}(\ln 2) \tag{20}$$

In the exact world this would be a daunting sum. In Asymptopia we will split the sum into the main terms and the small terms. Where to make the split is part of the *art* of Asymptopia which we discuss further below. For now, let u be the first index with $x_u \leq n \ln^{-2} n$. Applying (19) only down to x_{u-1} and adding we get

$$\pi(n) - \pi(x_u) \leq \sum_{i=0}^{u-1} \frac{x_i}{\ln(x_i/2)}(\ln 2) \tag{21}$$

Now we use the trivial bound $\pi(x_u) \leq x_u \leq n \ln^{-2} n$. While this is a "bad" bound for $\pi(x_u)$ it is a negligible value for us and

$$\pi(n) \leq o(\frac{n}{\ln n}) + \sum_{i=0}^{u-1} \frac{x_i}{\ln(x_i/2)}(\ln 2) \tag{22}$$

As x_i is decreasing so is the denominator $\ln(x_i/2)$ which pushes the sum (22) up. However, all terms in the sum have $x_i/2 > n \ln^{-2} n/2$. The ln function is going down, but not too far down. Each denominator

$$\ln(x_i/2) \geq \ln(n \ln^{-2} n/2) = \ln n - 2 \ln \ln n - \ln 2 = (1 - o(1)) \ln n \tag{23}$$

Thus

$$\sum_{i=0}^{u-1} \frac{x_i}{\ln(x_i/2)}(\ln 2) \leq \frac{1 + o(1)}{(\ln n)(\ln 2)} \sum_{i=0}^{u-1} x_i \tag{24}$$

Now $x_0 = n$ and $x_i \sim n2^{-i}$ (indeed, to be totally formal, $x_i \leq n2^{-i} + 1$) so that

$$\sum_{i=0}^{u-1} x_i \leq 2n(1 + o(1) \tag{25}$$

and (22) gives

$$\pi(n) \leq \frac{n}{\ln n}\frac{2}{\ln 2}(1 + o(1)) \tag{26}$$

`Selecting the Split:` When we chose u above there was a lot of room but still, care had to be taken. Knowing the answer in advance helps. Suppose we let u be the first index with $x_u < S$ and consider which values of S might work. It helps (as is frequently the case) to know [3] that $\pi(n) = \Theta(n/\ln n)$. In the argument we will be adding S and so we want $S = o(n/(lnn))$. But also the densities are going down in i when we look at $\pi(x_i) - \pi(x_{i+1})$ and we want them all to be $(1 + o(1))/(\ln n)$. As the last one will be $\sim 1/\ln(S)$ we will want $\ln(S) \sim \ln(n)$ which in turn requires $S = n^{1-o(1)}$. Indeed, any $S = n^{1-o(1)}$ with $S \ll (n/(\ln n)$ could have been used. Looking ahead at the argument we will be adding S. This leads us to require that $S = o(n/\ln n)$. Having finished the argument it is instructive to look back. The main intervals are roughly $[n, n/2), [n/2, n/4), \ldots$. In the first interval the upper bound for the density of primes from (19) is roughly $2/(\ln n)(\ln 2)$. This upper bound continues down to S, as $\ln(S) \sim \ln(n)$. Thus the upper bound on the total number of primes is at most S (which we choose to be negligible) plus what the number of primes would be if each interval had prime density $\frac{2}{\ln 2}\frac{1}{\ln n}$. The intervals total at most n values (actually a bit less since we cut it off at S) and so the main contribution to the prime count is $\sim \frac{2}{\ln n}\frac{n}{\ln n}$.

1.4 PMT with Constant

`Note:` This section gets quite technical and should be considered optional.

Here we show Theorem 1.4. That is, we *assume* that there is a constant c such that $\pi(n) \sim c(n/(\ln n)$ and then show that c must be 1. It is a big *if.* *A priori,* from Theorems 1.2,1.3 the ratio of $\pi(n)$ to $n/(\ln n)$ could oscillate between two positive constants, never approaching a limit.

We consider the factorization (11) more carefully. Our goal will be to show that if $c \neq 1$ then the left and right hand sides cannot match. We split the primes from 1 to n into intervals. We shall let K be a large but fixed

[3] Actually, a good hunch is useful. If the hunch turns out to be wrong the calculations will not come out as you wanted.

constant. (More about just how large later.) For $1 \le i < K$ let P_i denote the set of primes p with

$$\frac{n}{i+1} < p \le \frac{n}{i} \tag{27}$$

and let SP (small primes) denote the set of primes p with $p < \frac{n}{K}$. Let V_i, $1 \le i < K$ denote the contribution of the $p \in P_i$ to the factorization (11). That is, V_i is the product of $p_j^{\alpha_j}$ in (11), where p_j is restricted to P_i. Similarly let V_{SP} denote the contribution of the $p \in SP$ to the factorization (11). That is, V_i is the product of $p_j^{\alpha_j}$ in (11), where p_j is restricted to SP.

We first show that SP makes a relatively small contibution to (11). There are $\le \pi(n/K)$ primes $p \in SP$ and each (12) contributes at most a factor of n so that $V_{SP} \le n^{\pi(n/K)}$. From Theorem 1.3 gives $\pi(n/K) < (2\ln 2) + o(1))(n/K)/\ln(n/K)$. With K fixed, $\ln(n/K) \sim \ln(n)$ so that $\pi(n/K) < (\ln 2 + o(1))(n/K)/\ln(n)$. Thus (27),

$$V_{SP} < n^{(2\ln 2+o(1))(n/K)/\ln(n)} = 2^{(2n/K)(1+o(1))} \tag{28}$$

so that

$$\ln(V_{SP}) < \frac{2n\ln 2}{K}(1+o(1)) \tag{29}$$

While this is not a small number in absolute terms it will be relatively small compared to the total contribution which is $2^{n(1+o(1))}$.

For $1 \le i < K$ we now look at V_i. As all primes considered have $p > \frac{n}{K}$ and K is fixed they have $p > \sqrt{n}$. Thus the sum of Theorem 1.5 has only one term. Theorem 1.6 with $a = n/2$ is then simply

$$v_p(\binom{n}{n/2}) = \lfloor n/p \rfloor - 2\lfloor n/2p \rfloor \tag{30}$$

This is either zero or one and is one precisely when $\lfloor n/p \rfloor$ is odd. We have *designed* P_i so that $\lfloor n/p \rfloor = i$ for $p \in P_i$. When i is even no primes $p \in P_i$ appear in the factorization (11) (or, the same thing, they appear with exponent zero) and so $V_i = 1$. (For example, with $\frac{n}{7} < p \le \frac{n}{6}$, $n!$ has six factors of p and $(n/2)!^2$ has twice three factors of p and they all cancel.)

Now suppose $1 \le i < K$ is odd. Then V_i is simply the product of all primes $p \in P_i$. Each such prime p lies between $\frac{n}{K}$ and n and so can be considered $p = n^{1+o(1)}$. The number of such primes is $\pi(n/i) - \pi(n/(i+1))$. In this range $\ln(n/i) \sim \ln n$. Our assumption for Theorem ww3 then gives that $\pi(n/i) \sim c\frac{n}{i\ln n}$ and that that $\pi(n/(i+1) \sim c\frac{n}{(i+1)\ln n}$. We deduce that the number of primes is $\sim c\frac{n}{\ln n}(\frac{1}{i} - \frac{1}{i+1}$. (`Caution:` Subtraction in Asymptopia is dangerous! It is critical here that $i \le K$ and that K is a fixed constant, so $\frac{1}{i}$ and $\frac{1}{i+1}$ is a positive constant. Were, say, $K = \ln\ln n$ we could not do the subtraction. With $i \sim (\ln\ln n)/2$, for example, the asymptotics of $\pi(n/i)$ and $\pi(n/(i+1)$ would be the same and so one could *not* deduce the asymptotics of their difference!) Thus

$$V_i = n^{c(1+o(1))(n/(\ln n))(\frac{1}{i}-\frac{1}{i+1})} \tag{31}$$

and

$$\ln(V_i) \sim cn(\frac{1}{i} - \frac{1}{i+1}) \tag{32}$$

From the factorization (11) Then

$$\ln\left(\binom{n}{n/2}\right) = \ln V_{SP} + \sum \ln(V_i) \tag{33}$$

For convenience, assume $K = 2T$ is even so we can write the odd $i < K$ as $2j-1$, $1 \le j \le T$. From Chapter xxx, the left hand side is $\sim n\ln 2$. Thus

$$(1+o(1))n\ln 2 = cn(1+o(1))\sum)j = 1^T(\frac{1}{2j-1} - \frac{1}{2j}) + \ln V_{SP} \tag{34}$$

Dividing by n

$$(1+o(1))(\ln 2) = c(1+o(1))\sum_{k=1}^{2T-1} \frac{(-1)^{k+1}}{k} + \frac{1}{n}\ln V_{SP} \tag{35}$$

We need [4] the fact that

$$\ln 2 = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{i} = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \ldots \tag{36}$$

We can now see the idea. The $\ln(V_{SP})$ will be negligible and (35) becomes $\ln 2 = c(\ln 2)$. The actual argument consists of eliminating all $c \ne 1$.

Suppose $c > 1$. Select $K = 2T$ so that $c\sum_{k=1}^{2T-1} \frac{(-1)^{k+1}}{k} > \ln 2$. As $\ln V_{SP} \ge 0$ the right hand side of (35) would be bigger than the left hand side.

Suppose $c < 1$. Applying the upper bound (29), the right hand side of (35) would be at most $c\sum_{k=1}^{2T-1} \frac{(-1)^{k+1}}{k} + \frac{2\ln 2}{K}$. As $K \to \infty$, this sum approaches $c\ln 2$ which is less than $\ln 2$. Thus we may select K [5] so that

[4] Again, from Calculus!

[5] A subtle wrinkle here, while we examine behavior as $K \to \infty$ we select K a constant, dependent only on c.

this sum is less than $\ln 2$. But now the right hand side of (35) would be smaller than the left hand side.

Both assumptions led to a contradiction and since we *assummed* that c existed, it must be that $c = 1$.

1.5 Telescoping

Suppose we have a reasonable function $f(x)$ and we wish to asymptotically evaluate $\sum_{p \le n} f(p)$. We assume the Prime Number Theorem 1, giving the asymptotics of $\pi(s)$ as $s \to \infty$. On an intuitive level we think of $1 \le s \le n$ as being prime with "probability" $\pi(s)/s \sim 1/(\ln s)$. Then s, $1 \le s \le n$ would contribute $f(s)/(\ln s)$ to the sum and $\sum_{p \le n} f(p)$ would be roughly $\sum_{s \le n} f(s)/(\ln s)$. This is not a proof, integers are either prime or they aren't, yet surprisingly we can often get this intuitive result. The key is called telescoping. We write

$$\sum_{p \le n} f(p) = \sum_{s=2}^{n} f(s)(\pi(s) - \pi(s-1)) \tag{37}$$

Reversing sums (and noting $\pi(1) = 0$)

$$\sum_{s=2}^{n} f(s)(\pi(s) - \pi(s-1)) = f(n)\pi(n) + \sum_{s=2}^{n-1} \pi(s)(f(s) - f(s+1)) \tag{38}$$

While (38) its effectiveness depends on our ability to asymptotically calculate the sum. An important success is when $f(s) = \frac{1}{s}$, we ask for the asymptotics of

$$F(n) = \sum_{p \le n} \frac{1}{p} \tag{39}$$

The first term of (38) is then $\sim \frac{1}{n}\frac{n}{\ln n} = o(1)$. The sum is asymptotically $\sum \frac{s}{\ln s}\frac{1}{s(s+1)} \sim \sum \frac{1}{s \ln s}$, the sum from $s = 1$ to $n - 1$. From Chapter xxx,

$$\sum_{s=2}^{n-1} \frac{1}{s \ln s} \sin \int_1^n \frac{dx}{x \ln x} = \ln\ln n \tag{40}$$

That is, $F(n) \sim \ln\ln n$. For another example, take $f(s) = s$ so that $F(n) = \sum_{p \le n} p$. Then

$$F(n) = n\pi(n) - \sum_{s=2}^{n-1} \pi(s) \sim \frac{n^2}{\ln n} - \int_2^{n-1} \frac{s}{\ln s} ds \tag{41}$$

While the integrand cannot be precisely integrated we can handle it in Asymptopia. Our notion is that $\ln s \sim \ln n$ for "most" $2 \leq s \leq n-1$. We split the integral at some $n^{1-o(1)}$, let us take $u(n) = n \ln^{-10} n$ for definiteness. For $u(n) \leq s$, $\ln(s) \geq \ln n - 10 \ln\ln n \sim \ln n$ so that

$$\int_{u(n)}^{n-1} \frac{s}{\ln s} ds \sim \int_{u(n)}^{n-1} \frac{s}{\ln n} ds \sim \frac{n^2}{2 \ln n} \tag{42}$$

For $s \leq u(n)$ we bound $\frac{s}{\ln s} \leq s$ so that

$$\int_{2}^{u(n)} \frac{s}{\ln s} ds \leq \int_{0}^{u(n)} s ds \sim \frac{n^2}{2 \ln^{20} n} \tag{43}$$

As the upper bound (43) is $o(n^2/\ln n)$ it has a negligible effect and the total integral

$$\int_{2}^{n-1} \frac{s}{\ln s} ds \sim \frac{n^2}{2 \ln n} \tag{44}$$

Subtracting, (41) gives

$$\sum_{p \leq n} p \sim \frac{n^2}{\ln n} - \frac{n^2}{2 \ln n} \sim \frac{n^2}{2 \ln n} \tag{45}$$