Solution

Week 18 (2/13/03)

Distribution of primes

A necessary and sufficient condition for N to be prime is that N have no prime factors less than or equal to $\sqrt{N}$. Therefore, under the assumption that a prime p divides N with probability $1/p$, the probability that N is prime is

$$P(N) = \left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)\left(1-\frac{1}{5}\right)\left(1-\frac{1}{7}\right)\cdots\left(1-\frac{1}{p_{(\sqrt{N})}}\right), \tag{1}$$

where $p_{(\sqrt{N})}$ denotes the largest prime less than or equal to $\sqrt{N}$. Our strategy for solving for $P(N)$ will be to produce a differential equation for it.

Consider $P(N+n)$, where n is an integer that satisfies $\sqrt{N} \ll n \ll N$. We have

$$P(N+n) = \left(1-\frac{1}{2}\right)\left(1-\frac{1}{3}\right)\left(1-\frac{1}{5}\right)\left(1-\frac{1}{7}\right)\cdots\left(1-\frac{1}{p_{(\sqrt{N+n})}}\right), \tag{2}$$

where $p_{(\sqrt{N+n})}$ denotes the largest prime less than or equal to $\sqrt{N+n}$. Eq. (2) may be written as

$$P(N+n) = P(N)\left(1-\frac{1}{p_1}\right)\left(1-\frac{1}{p_2}\right)\cdots\left(1-\frac{1}{p^{(\sqrt{N+n})}}\right), \tag{3}$$

where the p_i are all the primes between $\sqrt{N}$ and $\sqrt{N+n}$. Let there be k of these primes. Since $n \ll N$, we have $\sqrt{N+n}/\sqrt{N} \approx 1$. Therefore, the p_i are multiplicatively all roughly the same. To a good approximation, we may therefore set them all equal to $\sqrt{N}$ in eq. (3). This gives

$$P(N+n) \approx P(N)\left(1-\frac{1}{\sqrt{N}}\right)^k. \tag{4}$$

We must now determine k. The number of numbers between $\sqrt{N}$ and $\sqrt{N+n}$ is

$$\begin{aligned}
\sqrt{N+n}-\sqrt{N} &= \sqrt{N}\sqrt{1+\frac{n}{N}}-\sqrt{N} \\
&\approx \sqrt{N}\left(1+\frac{n}{2N}\right)-\sqrt{N} \\
&= \frac{n}{2\sqrt{N}}\,.
\end{aligned} \tag{5}$$

Each of these numbers has roughly a $P(\sqrt{N})$ chance of being prime. Therefore, there are approximately

$$k \approx \frac{P(\sqrt{N})n}{2\sqrt{N}} \tag{6}$$

prime numbers between $\sqrt{N}$ and $\sqrt{N+n}$.

Since $n \ll N$, we see that $k \ll \sqrt{N}$. Therefore, we may approximate the $(1-1/\sqrt{N})^k$ term in eq. (4) by $1-k/\sqrt{N}$. Using the value of k from eq. (6), and writing $P(N+n) \approx P(N) + P'(N)n$, we can rewrite eq. (4) as

$$P(N) + P'(N)n \approx P(N)\left(1 - \frac{P(\sqrt{N})n}{2N}\right). \tag{7}$$

We therefore arrive at the differential equation,

$$P'(N) \approx -\frac{P(N)P(\sqrt{N})}{2N}\,. \tag{8}$$

It is easy to check that the solution for P is

$$P(N) \approx \frac{1}{\ln N}\,, \tag{9}$$

as we wanted to show.

Remarks:

1. It turns out (under the assumption that a prime p divides N with probability $1/p$) that the probability that N has exactly n prime factors is

$$P_n(N) \approx \frac{(\ln\ln N)^{n-1}}{(n-1)!\ln N}\,. \tag{10}$$

Our original problem dealt with the case $n = 1$, and eq. (10) does indeed reduce to eq. (9) when $n = 1$. Eq. (10) can be proved by induction on n, but the proof I have is rather messy. If anyone has a clean proof, let me know.

2. We should check that $P_1(N) + P_2(N) + P_3(N) + \cdots = 1$. The sum must equal 1, of course, because every number N has *some* number of divisors. Indeed (letting the sum go to infinity, with negligible error),

$$\begin{aligned} \sum_{n=1}^{\infty} P_n(N) &= \sum_{n=1}^{\infty} \frac{(\ln\ln N)^{n-1}}{(n-1)!\ln N} \\ &= \frac{1}{\ln N}\sum_{m=0}^{\infty}\frac{(\ln\ln N)^m}{m!} \\ &= \frac{e^{\ln\ln N}}{\ln N} \\ &= 1. \end{aligned} \tag{11}$$

3. We can also calculate the expected number, $\overline{n}$, of divisors of N. To do this, let's calculate $\overline{n-1}$ (which is a little cleaner), and then add 1.

$$\begin{aligned} \overline{n-1} &= \sum_{n=1}^{\infty}(n-1)P_n(N) \\ &\approx \sum_{n=2}^{\infty}\frac{(\ln\ln N)^{n-1}}{(n-2)!\ln N} \\ &= \frac{\ln\ln N}{\ln N}\sum_{k=0}^{\infty}\frac{(\ln\ln N)^k}{k!} \\ &= \ln\ln N. \end{aligned} \tag{12}$$

We can now add 1 to this to obtain $\overline{n}$. However, all our previous results have been calculated to leading order in N, so we have no right to now include an additive term of 1. To leading order in N, we therefore have

$$\overline{n} \approx \ln \ln N. \tag{13}$$

4. There is another way to calculate $\overline{n}$, without using eq. (10). Consider a group of M numbers, all approximately equal to N. The number of prime factors among all of these M numbers (which equals $M\overline{n}$ by definition) is given by[1]

$$M\overline{n} = \frac{M}{2} + \frac{M}{3} + \frac{M}{5} + \frac{M}{7} + \cdots. \tag{14}$$

   Since the primes in the denominators occur with frequency $1/\ln x$, this sum may be approximated by the integral,

$$M\overline{n} \approx M \int_1^N \frac{dx}{x \ln x} = M \ln \ln N. \tag{15}$$

   Hence, $\overline{n} \approx \ln \ln N$, in agreement with eq. (13).

5. For which n is $P_n(N)$ maximum? Since $P_{n+1}(N) = (\ln \ln N/n) P_n(N)$, we see that increasing n increases $P_n(N)$ if $n < \ln \ln N$. But increasing n decreases $P_n(N)$ if $n > \ln \ln N$. So the maximum $P_n(N)$ is obtained when

$$n \approx \ln \ln N. \tag{16}$$

6. The probability distribution in eq. (10) is a Poisson distribution, for which the results in the previous remarks are well known. A Poisson distribution is what arises in a random process such as throwing a large number of balls into a group of boxes. For the problem at hand, if we take $M(\ln \ln N)$ primes and throw them down onto M numbers (all approximately equal to N), then the distribution of primes (actually, the distribution of primes minus 1) will be (roughly) correct.

[1] We've counted multiple factors of the same prime only once. For example, we've counted 16 as having only one prime factor. To leading order in N, this method of counting gives the same $\overline{n}$ as assigning four prime factors to 16 gives (due to the fact that $\sum(1/p^k)$ converges for $k \geq 2$).