5. The Sieve of Eratosthenes

What is an efficient method to generate all of the primes up to x? The first observation is that if $m < x$ is composite then it must have a prime factor p at most $\sqrt{x}$.

So scratch out all multiples of 2, then 3, then 5, until the last integer remaining up to $\sqrt{x}$. The r initial integers $p_1, p_2, \ldots, p_r$ you scratch out are all prime and any integer remaining up to x is also prime.

To get good estimates for $\pi(x)$, one can modify this process a little bit. The idea is to make a more judicious choice of r. Pick r such that the first r primes $p_1, p_2, \ldots, p_r$ are all less than $\sqrt{x}$. We are going to choose the optimal value for r at the end. Note that the remaining primes are never multiples of $p_1, p_2, \ldots, p_r$, that is, the remaining primes belong to the set of integers which are not multiples of $p_1, p_2, \ldots, p_r$.

Let

$$P = \{\, n \in \mathbb{N} \,|\, 1 < n \leq x \text{ and } n \text{ is a not a multiple of } p_1, p_2, \ldots, p_r \,\},$$

so that P is the set of integers from 2 to x which are not multiples of $p_1, p_2, \ldots, p_r$. Let $A(x, r)$ be the cardinality of P.

If p is a prime from 1 to n then either p is one of $p_1, p_2, \ldots, p_r$ or p belongs to P. It follows that

$$\pi(x) \leq r + A(x, r).$$

We want to estimate $A(x, r)$. Let M_i be the set of integers from 1 to n which are multiples of p_i. Let M_{ij} be the set of integers from 1 to n which are multiples of both p_i and p_j. As p_i and p_j are coprime,

$$M_{ij} = M_i \cap M_j.$$

Note that

$$|M_i| = \llcorner \frac{x}{p_i} \lrcorner \qquad \text{and} \qquad |M_{ij}| = \llcorner \frac{x}{p_i p_j} \lrcorner,$$

and so on. It follows by inclusion-exclusion that

$$A(x, r) = \llcorner x \lrcorner - \sum_{i=1}^{r} \llcorner \frac{x}{p_i} \lrcorner + \sum_{i \neq j \leq r} \llcorner \frac{x}{p_i p_j} \lrcorner + \cdots + (-1)^r \llcorner \frac{x}{p_1 p_2 \ldots p_r} \lrcorner.$$

Suppose that we approximate the RHS by simply ignoring all of the round downs,

$$x - \sum_{i=1}^{r} \frac{x}{p_i} + \sum_{i \neq j \leq r} \frac{x}{p_i p_j} + \cdots + (-1)^r \frac{x}{p_1 p_2 \ldots p_r}.$$

The worse case scenario for the error is

$$1 + \binom{r}{1} + \binom{r}{2} + \cdots + \binom{r}{r} = 2^r.$$

It follows that

$$\pi(x) \leq r + x \prod_{i=1}^{r} \left(1 - \frac{1}{p_i}\right) + 2^r.$$

The key point in using this formula is to make a judicious choice of r. We want to choose r relatively small. For this we need a good estimate of the middle term.

Theorem 5.1. *If $x \geq 2$ then*

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right) < \frac{1}{\log x}.$$

Proof. We compute the product of the reciprocals,

$$\prod_{p \leq x} \frac{1}{1 - \frac{1}{p}} = \prod_{p \leq x} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \dots\right).$$

Consider what happens if we expand the RHS. If m is an integer which is a product of primes less than x then the term $\frac{1}{m}$ appears somewhere in the expansion of this product.

Now any integer $m \leq x$ is a product of primes less than x and so

$$\begin{aligned} \prod_{p \leq x} \frac{1}{1 - \frac{1}{p}} &> \sum_{k=1}^{n} \frac{1}{k} \\ &> \int_{1}^{\ulcorner x \urcorner} \frac{\mathrm{d}u}{u} \\ &> \log x. \end{aligned}$$

□

Theorem 5.2. *We have*

$$\pi(x) = O\left(\frac{x}{\log \log x}\right).$$

Proof.

$$\begin{aligned}
\pi(x) &\le r + x\prod_{i=1}^{r}\left(1-\frac{1}{p_i}\right) + 2^r \qquad \text{as proved above} \\
&\le 2^{r+1} + x\prod_{i=1}^{r}\left(1-\frac{1}{p_i}\right) \qquad \text{as } r \le 2^r \\
&\le 2^{r+1} + \frac{x}{\log p_r} \qquad \text{using (5.1)} \\
&\le 2^{r+1} + \frac{x}{\log r} \qquad \text{as } p_r \ge r \\
&\le 2^{\log x+2} + \frac{x}{\log\log x} \qquad \text{take } r = \ulcorner \log x \urcorner \\
&\le 4\cdot 2^{\log x} + \frac{x}{\log\log x} \\
&= O\left(2^{\log x}\right) + \frac{x}{\log\log x} \\
&\le o\left(\frac{x}{\log\log x}\right) + \frac{x}{\log\log x} \qquad \text{as } \log 2 < 1 \\
&= O\left(\frac{x}{\log\log x}\right).
\end{aligned}$$

□

It is interesting to note the strange connection between (5.2) and (5.1). We use (5.2) in the proof of (5.1). However, the two results have opposite conclusions.

(5.2) places an upper bound on the number of primes up x; there cannot be too many. By contrast, (5.1) places lower bounds on the number of primes up to x; there cannot be too few.

For example:

Lemma 5.3. *There are infinitely many n such that $p_n < n^2$.*

Proof. Suppose not; then there would be a natural number n_0 such that if $n > n_0$ then $p_n > n^2$. In this case

$$\begin{aligned}
\prod_{n=1}^{N}\left(1-\frac{1}{p_n}\right) &= M_0 \prod_{n=n_0}^{N}\left(1-\frac{1}{p_n}\right) \\
&\geq M_0 \prod_{n=n_0}^{N}\left(1-\frac{1}{n^2}\right) \\
&= M_0 \prod_{n=n_0}^{N}\frac{n-1}{n}\prod_{n=n_0}^{N}\frac{n+1}{n} \\
&= M_0 \frac{n_0-1}{N}\frac{N+1}{n_0} \\
&= M_0 \frac{n_0-1}{n_0}\frac{N+1}{N} \\
&\geq \frac{M_0}{2}.
\end{aligned}$$

But this contradicts the fact the product is supposed to go to zero by (5.1). □