Prime number heuristics

Primes by chance?

Torsten Ekedahl

February 21, 2005

The grand plan

1. Probable primes
   - The prime number theorem
   - Twinning primes
   - Sifting primes
1. Prime renormalisation
   - Twins revisited
   - Another go
   - Pitfalls
1. Cramér's model

Number of primes

- $\pi(x)$ number of primes $\leq x$
- Prime number theorem
$$\pi(x) \sim \frac{x}{\log x}$$
- π grows slowly $\rightsquigarrow$
$$\pi(x) - \pi(y) \approx \frac{x-y}{\log x}$$
- Heuristic (=crazy) interpretation:
$$\text{Prob}(n \text{ is prime}) = \frac{1}{\log n}$$
- Expected number of primes $\leq x$:
$$\sum_{2 \leq n \leq x} \frac{1}{\log n}$$

Reality check

Compare this prediction with $\pi(x)$ och $x/\log x$:

n	$\pi(10^n)$	$10^n/\log 10^n$	Expectation
1	4	4	6
2	25	22	30
3	168	145	178
4	1229	1086	1246
5	9592	8686	9630
6	78498	72382	78628
7	664579	620421	664918
8	5761455	5428681	5762209
9	50847534	48254942	50849235
10	455052511	434294482	455055615

Approx. half the numbers right; probabilistically reasonable.

Twin primes

n and $n+2$ are primes

$$\text{Prob}(n \text{ and } n+2 \text{ prime}) = \frac{1}{\log n} \cdot \frac{1}{\log(n+2)}$$

x	Twins	Expect. value	Quotient
10	2	3.4038	0.5876
100	8	9.8001	0.8163
1000	35	34.1945	1.0236
10000	205	161.7370	1.2675
100000	1224	945.2490	1.2949
1000000	8169	6246.4600	1.3078
10000000	58980	44499.0000	1.3254
15000000	83660	63241.4000	1.3229

Wrong limit!

The sieve of Eratosthenes

- Single prime divisibility

$$\mathrm{Prob}(p \nmid n) = 1 - \frac{1}{p}$$

- Chinese remainder theorem $\rightsquigarrow$

$$\mathrm{Prob}(p \nmid n \wedge q \nmid n) = \mathrm{Prob}(p \nmid n)\,\mathrm{Prob}(q \nmid n)$$

- Sift repeatedly:

$$\mathrm{Prob}(n \text{ is prime}) = \prod_{p \leq n} \left(1 - \frac{1}{p}\right)$$

- Elementary fact:

$$\prod_{p \leq n} \left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log n}, \gamma = 0.577216\ldots$$

$$\text{Prob}(n \text{ is prime}) \sim \frac{e^{-\gamma}}{\log n}$$

$$e^{-\gamma} \approx 0.561459 \neq 1$$

- Modify:

$$\text{Prob}(n \text{ is prime}) = \prod_{p \leq \sqrt{n}} \left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log \sqrt{n}} = \frac{2e^{-\gamma}}{\log n}$$

- $2e^{-\gamma} \approx 1.12292$. Close but no cigar!
- Error term:

$$\text{Prob}(p \not| n) = 1 - \frac{1}{p} + \mathcal{O}(\frac{p}{n})$$

Heuristic assumptions

- $2e^{-\gamma}$ because of "interval tails".
- The needed corrections are the same for all our problems.
- We can "renormalise" (give ∞/∞ a specific value).

$$\text{Prob}(n \text{ and } n+2 \text{ primes}) = \frac{\text{Prob}(n \text{ and } n+2 \text{ primes})}{\text{Prob}(m \text{ and } n \text{ prime})}\text{Prob}(m \text{ and } n \text{ prime})$$

m and n random with $m \approx n \rightsquigarrow$

$$\text{Prob}(m \text{ and } n \text{ prime}) = \frac{1}{\log^2 n}$$

$$\text{Prob}(n \text{ and } n+2 \text{ primes}) = \frac{\text{Prob}(n \text{ and } n+2 \text{ primes})}{\text{Prob}(m \text{ and } n \text{ prime})}\frac{1}{\log^2 n}$$

"Tail errors" will (hopefully) cancel.

$$\mathrm{Prob}(m, m' \text{ primes}) = \prod_{p<n} \left(1 - \frac{1}{p}\right)^2$$

$$\mathrm{Prob}(n, n+2 \text{ primes}) = \frac{1}{2} \prod_{2<p<n} \left(1 - \frac{2}{p}\right)$$

Not independent!

$$\frac{\mathrm{Prob}(n, n+2 \text{ primes})}{\mathrm{Prob}(m, m' \text{ primes})} = 2 \prod_{2<p<n} \frac{1 - \frac{2}{p}}{\left(1 - \frac{1}{p}\right)^2}$$

Conclusion

$$\frac{1-\frac{2}{p}}{(1-\frac{1}{p})^2} = 1 + O\left(\frac{1}{p^2}\right)$$

$\rightsquigarrow$

$$\prod_{2<p} \frac{1-\frac{2}{p}}{(1-\frac{1}{p})^2} \quad \text{is convergent}$$

Renormalisation!

$\rightsquigarrow$

$$\pi_{twin}(x) \sim 2 \prod_{2<p} \frac{1-\frac{2}{p}}{(1-\frac{1}{p})^2} \frac{x}{\log^2 x}$$

Reality check

x	Twins	Expect. value	Quotient
10	2	4.4942	0.4450
100	8	12.9393	0.6183
1000	35	45.1478	0.7752
10000	205	213.5452	0.9600
100000	1224	1248.0346	0.9807
1000000	8169	8247.3488	0.9905
10000000	58980	58753.0813	1.0039
15000000	83660	83499.1149	1.0019

The problem

How many prime numbers of the form $n^2 + 1$ are there?

Tentative answer

With $m \approx n^2 + 1$:

$$\frac{\text{Prob}(n^2+1 \text{ prime})}{\text{Prob}(m \text{ prime})} \text{Prob}(m \text{ prime})$$

$$\text{Prob}(m \text{ prime}) = \frac{1}{\log m} = \frac{1}{\log(n^2+1)} \approx \frac{1}{2\log n}$$

Easy part:

$$\mathrm{Prob}(m \text{ prime}) = \prod_{p<m}\left(1-\frac{1}{p}\right)$$

$\mathrm{Prob}(n^2+1 \text{ prime})$

- $p | n^2+1 \iff n^2 \equiv -1 \bmod p$
- $\exists n : n^2 \equiv -1 \bmod p \implies (-1)^{(p-1)/2} \equiv n^{p-1} \equiv 1$
- $\rightsquigarrow p \equiv 1 \bmod 4$
- Converse also true

Conclusion

- $p \equiv 1 \bmod 4$: $\mathrm{Prob}(p \nmid n^2+1) = 1 - 2/p$ (2 sol's)
- $p \equiv -1 \bmod 4$: $\mathrm{Prob}(p \nmid n^2+1) = 1 - 0/p$ (No sol's)

$\rightsquigarrow$

$$\mathrm{Prob}(n^2+1 \text{ prime}) = \prod_{p<n^2+1} \left(1 - \frac{c_p}{p}\right)$$

$$c_p = \begin{cases} 2 & \text{if } p \equiv 1 \bmod 4, \\ 0 & \text{if } p \equiv -1 \bmod 4. \end{cases}$$

$$\frac{\mathrm{Prob}(n^2+1 \text{ prime})}{\mathrm{Prob}(m \text{ prime})} = \prod_{2<p} \frac{1-\frac{c_p}{p}}{1-\frac{1}{p}} \quad \text{is convergent}$$

Final formula

$$\pi_{n^2+1}(x) \sim \prod_{2<p} \frac{1-\frac{c_p}{p}}{1-\frac{1}{p}} \frac{1}{\log x}$$

Reality check

$$\prod_{2<p} \frac{1 - \frac{c_p}{p}}{1 - \frac{1}{p}} \approx 1.37278$$

n	π_{n^2+1}	Expect. value	Quotient
100	19	20.4147	0.930703
1000	112	121.727	0.920095
10000	841	855.906	0.982584
100000	6656	6616.37	1.00599
1000000	54110	54025.2	1.00157

On the wrong path

How many prime numbers of the form $2^m \pm 1$ are there?

Tentative answer

- $\text{Prob}(2^m \pm 1 \text{ prime}) = 1/\log(2^m \pm 1)$.
- Expected number of primes $\leq n$:

$$\sum_{1<m\leq n} \frac{1}{2^m \pm 1} \sim \frac{\log n}{\log 2}$$

- Hence the expected number in all is infinite.

This gives the wrong answer.

Back to basics

- m odd $\implies 3|2^m+1$.
- Hence only $2^{2^k}+1$ can be prime.

$$\sum_k \frac{1}{\log(2^{2^k}+1)} < \infty$$

- $m = k\cdot\ell \implies 2^k-1|2^m-1$.
- Hence only 2^p-1 can be prime.

$$\sum_{p\leq n} \frac{1}{\log(2^p-1)} \approx \frac{1}{\log 2}\sum_{p\leq n}\frac{1}{p} \sim \frac{\log\log n}{\log 2}$$

Question

Can this be seen by sifting?

The model

- $z_n \in \{0, 1\}$ "Independent random variables"

$$\mathrm{Prob}(v_n = 1) = \frac{1}{\log n}$$

- $\Pi(x) := \sum_{n \leq x} z_n$ number of "random primes"
- Expectation value

$$E(\Pi(x)) = \sum_{n \leq x} \frac{1}{\log n} \sim \frac{x}{\log x}$$

Primes in small intervals

For $N > 2$

$$E(\Pi(x + \log^N x) - \Pi(x)) \sim \log^{N-1} x$$

Theorem of Maier (1985)

Contradicting Cramér's model!

Exists $\delta_N > 0$ s.t. for arbitrarily large x_+, x_-

$$\begin{aligned} \pi(x_+ + \log^N x_+) - \pi(x_+) &\geq (1 + \delta_N) \log^{N-1} x_+ \\ \pi(x_- + \log^N x_-) - \pi(x_+) &\leq (1 - \delta_N) \log^{N-1} x_- \end{aligned}$$

Maier’s argument

- Sift for small primes up to $\log x$. (Very small tails.)
- Then use Cramér’s model. (Using the prime number theorem to justify.)

Conclusion

Sifting is fundamental for primes.

More differences

Prime gap distribution

- Cramér's model (p_n, n'th prime):

$$\max_{p_n \leq x}(p_{n+1} - p_n) \sim \log^2 x$$

- Maier's model (sifting $+$ Cramér):

$$\max_{p_n \leq x}(p_{n+1} - p_n) \gtrsim 2e^{-\gamma} \log^2 x$$

- Numerical data not enough to distinguish between them.

In defence of Cramér

- Cramér’s model seems to work for small and large intervals. Maier’s counter example deals with medium sized intervals.
- Heuristics must be used sensibly... The examples discussed previously do not need the precision where Maier’s counter example lives.