THE TWIN PRIME CONJECTURE

QIAO ZHANG

1. Primer on Primes

Examples of primes are $2, 3, 5, 7, 11$. (Not 57 — Grothendieck said this was prime once). The fundamental theorem of arithmetic states that every positive integer has a unique representation as a product of primes. The analytic form of this theorem is

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1},$$

which is the Riemann zeta function $\zeta(s)$.
Let

$$\pi(x) = \#\{p \leq x\}.$$

Euclid showed that $\pi(x) \to \infty$. In fact, he showed $\pi(x) > \log_2(\log_2(x))$. For example, $\pi\left(10^6\right) > 4.13$.
Legendre conjectured in 1798/1802(?) that

$$\pi(x) \sim \frac{x}{\log x - 1.08366...}$$

Gauss conjectured that

$$\pi(x) \sim \int_2^x \frac{dt}{\log t}.$$

Can we compare these results? Integrate by parts to get

$$\begin{aligned} \pi(x) &\sim \int_2^x \frac{dt}{\log t} \sim \frac{x}{\log x} + \frac{x}{(\log x)^2} + ... = \frac{x}{\log x}\left(1 + \frac{1}{\log x} + ...\right) \\ &\sim \frac{x}{\log x - 1}. \end{aligned}$$

Who is closer? The integral is definitely best. For example, for $x = 10^{10}$. We have $\pi\left(10^{10}\right) = 455,052,511$.
Legendre gives us:

$$455,743,004$$

Gauss gives us (from $\frac{x}{\log x - 1}$).

$$454,011,971$$

The Gauss integral gives

$$455,055,613.$$

In 1859, Riemann (student of Gauss) published his only number theory paper with the Riemann hypothesis conjectures and showed that the prime number theorem is related to the zeta function. In 1896, Hadamard and de la Poussin showed

$$\pi(x) \sim li(x).$$

The world record for today is

$$\pi(x) \sim li(x) + \mathcal{O}\left(x \exp\left(-c\frac{(\log x)^{3/5}}{(\log\log x)^{1/5}}\right)\right).$$

The Riemann hypothesis is equivalent to

$$\pi(x) \sim li(x) + \mathcal{O}\left(\sqrt{x}(\log x)^2\right).$$

Date: October 2, 2013.

Numerical data has shown that $\pi(x) < li(x)$. In 1914, Littlewood proved that the inequality is for all x, and in fact that it switches back and forth infinitely often. His student Skewes showed that there exists an example x such that $\pi(x) \geq li(x)$ with

$$2 < x < 10^{10^{10^{963}}}$$

The current record is

$$2 < x < 10^{311}.$$

(No example is known.)
Dirichlet considered how prime numbers are distributed in certain arithmetic progressions:

$$\pi(x; q, a) = \#\{p \leq x : p \equiv a \bmod q\} \sim \frac{\pi(x)}{\varphi(q)}.$$

Vinogradov showed (1965) that

$$\pi(x; q, a) \sim \frac{li(x)}{\varphi(q)} + \mathcal{O}_q\left(x \exp\left(-c\frac{(\log x)^{3/5}}{(\log\log x)^{1/5}}\right)\right).$$

The generalized Riemann hypothesis (GRH) implies that

$$\pi(x; q, a) \sim \frac{li(x)}{\varphi(q)} + \mathcal{O}_q\left(\sqrt{x}(\log x)^2\right). \tag{1}$$

The GRH involves

$$L(s, \chi) = \sum \frac{\chi(n)}{n^s} = \prod \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}.$$

The GRH is that for all $\chi \,(\mathrm{mod}\, q)$, the nontrivial zeros of $L(s, \chi)$ are all on the critical line $\mathrm{Re}(s) = \frac{1}{2}$.
Page showed that if $q \leq (\log x)^2$, then (BestError) is true.
Siegel showed that if $q \leq (\log x)^A$, then (BestError) is true for all $A > 0$. Because of the proof method of his, there is no control of the constant in the error term.
Note the GRH implies that the best error estimate is true for $q \leq x^{1/2+\varepsilon}$ for all $\varepsilon > 0$.
Bombieri-Vinogradov (1960s-1970s) showed this is true on average. They proved that

$$\sum_{q \leq Q} \max_{y \leq x} \max_{(a,q)=1} \left|\pi(x; q, a) - \frac{li(x)}{\varphi(q)}\right| \leq \frac{x}{(\log x)^A},$$

where $Q = \frac{\sqrt{x}}{(\log x)^B}$.
In the early 1980s, Fouvry-Friedlander-Iwaniec showed that at some cost (no absolute values, only certain categories of arithmetic progressions), we get Q up to $x^{4/7}$.
Then Elliott-Holberstam conjectured that $Q = x^{1-\varepsilon}$ is possible.
(Some of these results are related to Yitang Zhang's work.)
The prime number theorem tells us that

$$\pi(x) \sim \frac{x}{\log(x)},$$

so the average gap between primes is $\log(x)$. The question is the variance.

Conjecture 1. *There are infinitely many twin primes. More precisely,*

$$\#\{p \leq x : p, p+2 \text{ prime}\} \sim \mathfrak{S}(\{0,2\}) \frac{x}{(\log x)^2},$$

where

$$\begin{aligned} \mathfrak{S}(\{0,2\}) &= 2 \prod_{\ell \geq 3,\ \ell \text{ prime}} \left(1 - \frac{2}{\ell}\right)\left(1 - \frac{1}{\ell}\right)^{-2} \\ &= 1.3... \end{aligned}$$

2. Lecture II

We know that

$$\pi(x) \sim \frac{x}{\log x},$$

so the average prime gap is $\log(x)$. But what is the variance? Probability distributions for gaps? Recall the conjecture above. In fact, we have

Conjecture 2. *Let $\mathcal{H} \subseteq \mathbb{Z}_+$ be a finite subset, and for every prime ℓ we write*

$$\nu_{\mathcal{H}}(\ell) = \#\text{ residue classes mod } \ell \text{ occupied by the elements in } \mathcal{H}.$$

Let $k = |\mathcal{H}|$. Then

$$\begin{aligned} & \#\{n \le x : n+h_1, ..., n+h_k \text{ all primes}\} \\ \sim\ & \mathfrak{S}(\mathcal{H}) \frac{x}{(\log x)^k}. \end{aligned}$$

where

$$\mathfrak{S}(\mathcal{H}) = \prod_{p \ge 2,\ prime} \left(1 - \frac{\nu_{\mathcal{H}}(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k}.$$

The twin prime conjecture is the case where $\mathcal{H} = \{0, 2\}$. For example, $\mathcal{H} = \{0, 1\}$ implies $\nu_{\mathcal{H}}(2) = 2$, and this implies $\mathfrak{S}(\mathcal{H}) = 0$. If the result of $\mathfrak{S}(\mathcal{H})$ is zero, we say $\mathcal{H}$ is admissible.
Where does this come from. There are a number of heuristic arguments for this. The first is called **Cramer's Model**, a probability argument. Consider a sequence of random variables $\{X(n)\}$ so that it sort of behaves like $X(n)$ is 1 if n is prime, 0 otherwise. Then

$$\begin{aligned} P(X(n) = 1) &= \frac{1}{\log n}, \\ P(X(n) = 0) &= 1 - \frac{1}{\log n}. \end{aligned}$$

How can we use this model? For example, if we wish to study the distribution of the prime gaps,

$$\begin{aligned} P(\text{the prime gap is } h) \text{ ``} &= \text{''} \\ & P(X(n+1) = 0, X(n+2) = 0, ..., X(n+h-1) = 0, X(n+h) = 1 : X(n) = 1) \\ &= P(X(n+1) = 0) P(X(n+2) = 0) ... \\ &= \left(1 - \frac{1}{\log(n+1)}\right)\left(1 - \frac{1}{\log(n+2)}\right) ... \left(1 - \frac{1}{\log(n+h-1)}\right) \frac{1}{\log(n+h)} \end{aligned}$$

assuming all events are independent.

$$\text{``}=\text{''} \left(1 - \frac{1}{\log n}\right)^{h-1} \frac{1}{\log n}.$$

Now, we sum up all the probabilities for $\alpha \log n < h < \beta \log n$: (where p' is the next prime after p)

$$\begin{aligned} & P(\alpha \log p < p' - p < \beta \log p) \\ \text{``} = \text{''} \quad & \sum_h P(.... : X(n) = 1) \\ \text{``} = \text{''} \quad & \sum_{\alpha \log p < h < \beta \log p} \left(1 - \frac{1}{\log p}\right)^{h-1} \frac{1}{\log p}. \end{aligned}$$

Recall, for $x \to 0$, $1 - x \sim e^{-x}$, so we get

$$\begin{aligned} \text{“} =\text{”} & \sum_{\alpha \log p < h < \beta \log p} \exp\left(-\frac{h-1}{\log p}\right) \frac{1}{\log p} \\ = & \sum_{\alpha < \frac{h}{\log p} < \beta} \exp\left(-\frac{h-1}{\log p}\right) \frac{1}{\log p} \\ \to & \int_\alpha^\beta e^{-t} dt. \end{aligned}$$

This suggests (Poisson distribution) that

$$P\left(\exists k \text{ primes in } [n, n + \log n]\right) \text{“=”} \frac{\mu^k}{k!} \exp(-\mu) = \frac{1}{k!e},$$

where

$$\mu = 1,$$

the average gap. What does this say for twin primes?

$$\begin{aligned} P(\text{twin primes}) &= P\left(X(n) = 1, X(n+1) = 0, X(n+2) = 1\right) \\ &= \frac{1}{\log n}\left(1 - \frac{1}{\log n}\right)\frac{1}{\log n} \sim \frac{1}{\log(n)^2}. \end{aligned}$$

This suggests that the number $\#\{\text{twin primes smaller than } x\} \sim \frac{x}{\log(x)^2}$.
However, this is wrong. Instead, we make corrections suggested by Hardy and Littlewood.
At a prime ℓ, originally we included a local factor $\left(1 - \frac{1}{\ell}\right)^2$. Now we take into account that they are very close. The correct local factor should be $1 - \frac{2}{\ell}$. So we need a local correction factor of $\left(1 - \frac{2}{\ell}\right)\left(1 - \frac{1}{\ell}\right)^{-2}$ for $\ell \geq 3$. So for $\ell = 2$, correction factor is $\left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{2}\right)^{-2} = 2$. This gives us

$$\#\{\text{twin primes } \leq x\} \sim \mathfrak{S}(\{0, 2\}) \frac{x}{(\log x)^2}.$$

Another heuristic argument is the Circle Method. Originally it was introduced as a bunch of power series, integrated over the unit circle. This is very complicated. Then this was reformulated in terms of exponential sums. If $e(x) = \exp(2\pi i x)$

$$\int_0^1 e(n\alpha)\, d\alpha = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\#\{p \leq x \text{ prime } : p + 2 \leq x \text{ prime}\} = \int_0^1 \left(\sum_{p \leq x, \text{ prime}} e(p\alpha)\right)\left(\sum_{p \leq x, \text{ prime}} e(-p\alpha)\right) e(2\alpha)\, d\alpha.$$

Then the number theory reduces to a calculus problem. Notice that

$$= \int_0^1 \left|\sum_{p \leq x, \text{ prime}} e(p\alpha)\right|^2 e(2\alpha)\, d\alpha.$$

We do not know how to estimate this. Hardy-Littlewood-Ramanujan in the 1930s observed that

$$\sum_{p \leq x, \text{ prime}} e(p\alpha)$$

is large when α is close to a rational number with a small denominator. It is small otherwise. It suggests that we can cut the integral into pieces.