15. Bruns Theorem

Definition 15.1. *Primes p and $p < q$ are called **twin primes** if $q = p + 2$.*

$\pi_2(x)$ is the number of pairs of twin primes up to x.

Conjecture 15.2. *There are infinitely many twin primes.*

Theorem 15.3.

$$\pi_2(x) \ll P(x) = \frac{x(\log\log x)^2}{\log^2 x}.$$

In particular, the sum of the reciprocals of all twin primes

$$\left(\frac{1}{3}+\frac{1}{5}\right)+\left(\frac{1}{5}+\frac{1}{7}\right)+\left(\frac{1}{11}+\frac{1}{13}\right)+\dots$$

is convergent.

(15.3) indicates why the twin prime conjecture is so hard. Indeed, even if there are infinitely many twin primes the sum of the reciprocals converges and so we cannot prove the existence of infinitely many twin primes a la Dirichlet.

Proof of (15.3). The last statement follows from the first by an easy application of the partial summation formula.

Consider integers of the form $a_n = n(n+2)$, $1 \leq n \leq x$. Remove those divisible by a prime $\leq y$, for some integer $y \leq \sqrt{x}$. Let $A(x, y)$ denote the number of remaining a_n. We have

$$\pi_2(x) \leq \pi(y) + A(x, y).$$

Our goal is to show that $A(x, y) \ll P(x)$ for some $y = y(x)$ such that $\pi(y) \ll P(x)$.

Put

$$R = \prod_{p \leq y} p.$$

We calculate using inclusion-exclusion

$$\begin{aligned} A(x,y) &= \lfloor x \rfloor - \sum_{p|R} \sum_{n \leq x, p|a_n} 1 + \sum_{p_1<p_2, p_1p_2|R} \sum_{n \leq x, p_1p_2|a_n} 1 - \dots \\ &= S_0 - S_1 + S_2 - \dots. \end{aligned}$$

One possibility is to proceed as we did in Lecture 5. However this introduces error terms which are too large. We introduce a variation on this theme:

Claim 15.4. *There is an inequality*

$$A(x,y) \leq S_0 - S_1 + S_2 - \cdots + S_{2k},$$

which is valid for every even index $0 < 2k \leq \pi(y)$.

Proof of (15.4). Suppose that a_n is divisible by exactly m prime factors of R. If $m = 0$ then a_n is counted once on the RHS and this is correct.

If $m > 0$ it is counted

$$C(a_n) = 1 - \binom{m}{1} + \binom{m}{2} - \cdots + \binom{m}{2k}$$

times on the RHS and not at all on the LHS and so we just need to check that $C(a_n) \geq 0$.

There are three cases. If $2k \leq 1/2(m+1)$ then we use the fact that the binomial coefficients

$$\binom{m}{l}$$

increase until the middle term (or pair of middle terms):

$$C(a_n) = 1 + \left\{\binom{m}{2} - \binom{m}{1}\right\} + \left\{\binom{m}{4} - \binom{m}{3}\right\} + \cdots > 0.$$

If $2k \geq m$ then

$$C(a_n) = (1-1)^m = 0.$$

Finally if $1/2(m+1) < 2k < m$ then we use the fact that the binomial coefficients decrease after the middle term:

$$\begin{aligned} C(a_n) &= (1-1)^n - \left\{-\binom{m}{2k+1} + \binom{m}{2k+2} - \dots\right\} \\ &= \left\{\binom{m}{2k+1} - \binom{m}{2k+2}\right\} + \dots \\ &> 0. \end{aligned}$$

□

Now we have (15.4) we still have to give a good approximation of the individual terms of the sum. For this we need to estimate the inner sums in the double summation for S_m,

$$T_d = \sum_{n \leq x, d | a_n} 1.$$

Note that

$$d = p_1 p_2 \dots p_l | R,$$

and $\mu(d) \neq 0$.

First suppose that d is odd. Then $n(n+2) \equiv 0 \mod d$ if and only if there is a factorisation $d = d_1 d_2$ such that

$$n \equiv 0 \mod d_1 \qquad \text{and} \qquad n + 2 \equiv 0 \mod d_2.$$

Claim 15.5. *n is determined, modulo d, by the factorisation.*

Proof of (15.5). Suppose also that

$$n \equiv 0 \mod e_1 \qquad \text{and} \qquad n+2 \equiv 0 \mod e_2,$$

where $d = e_1 e_2$.

It follows that the equations

$$\begin{aligned} x &\equiv 2 \mod d_1 \\ x &\equiv 0 \mod d_2 \\ x &\equiv 2 \mod e_1 \\ x &\equiv 0 \mod e_2 \end{aligned}$$

have a simultaneous solution, $x = n + 2$. Then by a result in 104A (concerning simultaneous solutions to systems of linear congruences) we must have

$$(d_1, e_2)|2 \qquad \text{and} \qquad (d_2, e_1)|2.$$

As d is odd this implies that both d_1 and e_2, and d_2 and e_1 are coprime. As d_1 divides $d = d_1 d_2$ it divides the product $e_1 e_2 = d$. As d_1 is coprime to e_2, d_1 divides e_1. By symmetry, it follows that $d_1 = e_1$ and $d_2 = e_2$, so that the factorisation is unique. □

It follows that there are exactly $\tau(d)$ solutions, modulo d, of $n(n+2) \equiv 0 \mod d$. Hence the number of $n \le x$ satisfying the congruence is

$$T_d = \lfloor \frac{x}{d} \rfloor \tau(d) + \theta_1 \tau(d) = \frac{x}{d}\tau(d) + \theta\tau(d)$$

where $0 \le \theta_1 \le 1$ so that $|\theta| \le 1$.

Now suppose that d is even. Then $n = 2m$ is even. In this case we just have to count the number of $m \le \frac{1}{2}x$ such that

$$m(m+1) \equiv 0 \mod \frac{1}{2}d.$$

By the same argument,

$$T_d = \frac{x/2}{d/2}\tau(d/2) + \theta\tau(d/2),$$

where $|\theta| \le 1$. If we define an auxiliary function

$$\tau'(d) = \begin{cases} \tau(d) & \text{if } d \text{ is odd} \\ \tau(d/2) & \text{if } d \text{ is even} \end{cases}$$

then we see that

$$T_d = \frac{x}{d}\tau'(d) + \theta\tau'(d) \qquad \text{where} \qquad |\theta| \le 1.$$

It follows that

$$A(x,y) \le x\left(1 - \sum_{p|R}\frac{\tau'(p)}{p} + \sum_{p_1p_2|R}\frac{\tau'(p_1p_2)}{p_1p_2} + \cdots + \sum_{p_1p_2\ldots p_{2k}|R}\frac{\tau'(p_1p_2\ldots p_{2k})}{p_1p_2\ldots p_{2k}}\right)$$

$$+\left(1 + \sum_{p|R}\tau'(p) + \sum_{p_1p_2|R}\tau'(p_1p_2) + \cdots + \sum_{p_1p_2\ldots p_{2k}|R}\tau'(p_1p_2\ldots p_{2k})\right).$$

We adopt the convention that the sums run over increasing primes, $p_1 < p_2 < \ldots$. If we fix l then

$$\begin{aligned}\sum_{p_1p_2\ldots p_l|R}\tau'(p_1p_2\ldots p_l) &\le 2^l\binom{\pi(y)}{l}\\ &= 2^l\frac{\pi(y)(\pi(y)-1)\ldots(\pi(y)-l+1)}{l!}\\ &= 2^l\frac{\pi^l(y)}{l!}.\end{aligned}$$

Thus the second term above, the error term, is at most

$$\begin{aligned}\sum_{p|R}\tau'(p) + \sum_{p_1p_2|R}\tau'(p_1p_2) + \cdots + \sum_{p_1p_2\ldots p_{2k}|R}\tau'(p_1p_2\ldots p_{2k}) &\le \sum_{l=1}^{2k}\pi^l(y)\frac{2^l}{l!}\\ &< \pi^{2k}(y)\sum_{l=1}^{2k}\frac{2^l}{l!}\\ &< \pi^{2k}(y)e^2.\end{aligned}$$

Now we turn to estimating the main term, the term involving x. As $\tau'(n)$ is multiplicative if we multiply out

$$\prod_{p|R}\left(1 - \frac{\tau'(p)}{p}\right) = \frac{1}{2}\prod_{p|R,p>2}\left(1 - \frac{2}{p}\right),$$

then we get the coefficient of x if we include

$$V_k = -\sum_{p_1p_2\ldots p_{2k+1}|R}\frac{\tau'(p_1p_2\ldots p_{2k+1})}{p_1p_2\ldots p_{2k+1}} + \sum_{p_1p_2\ldots p_{2k+2}|R}\frac{\tau'(p_1p_2\ldots p_{2k+2})}{p_1p_2\ldots p_{2k+2}} + \ldots.$$

Thus

$$A(x,y) \ll \frac{1}{2}\prod_{p|R,p>2}\left(1 - \frac{2}{p}\right) + \pi^{2k}(y) + x|V_k|.$$

Claim 15.6. *There is a constant $c \geq 0$ such that*

$$|V_k| < \sum_{l>2k} \left(\frac{2e \log\log y + ec}{l} \right)^l .$$

Proof of (15.6). Note that for $l \leq \pi(y)$ we have

$$\begin{aligned} \sum_{p_1 p_2 \dots p_l | R} \frac{\tau'(p_1 p_2 \dots p_l)}{p_1 p_2 \dots p_l} &\leq \sum_{p_1 p_2 \dots p_l | R} \frac{2^l}{p_1 p_2 \dots p_l} \\ &\leq \frac{1}{l!} \left(\sum_{p \leq y} \frac{2}{p} \right)^l . \end{aligned}$$

To see the last inequality, note that if we expand the multinomial

$$(t_1 + t_2 + \dots + t_s)^l$$

each of the

$$\binom{s}{l} \qquad \text{products} \qquad t_{h_1} t_{h_2} \dots t_{h_l},$$

$1 \leq h_1 < h_2 < \dots < t_{h_l} \leq s$, occurs with coefficient $l!$ (once can also prove the inequality directly by induction).

Therefore, by (8.2), we have

$$\begin{aligned} |V_k| &< \sum_{2k \leq l \leq \pi(y)} \frac{1}{l!} \left(\sum_{p \leq y} \frac{2}{p} \right)^l \\ &< \sum_{2k \leq l \leq \pi(y)} \frac{1}{l!} \left(2 \log\log y + c \right)^l , \end{aligned}$$

for some $c > 0$.

As

$$\begin{aligned} e^l &= \sum_{u=0}^{\infty} \frac{l^u}{u!} \\ &> \frac{l^l}{l!}, \end{aligned}$$

it follows that

$$l! > \left(\frac{l}{e} \right)^l .$$

Putting all of this together, we get

$$|V_k| < \sum_{l>2k} \left(\frac{2e \log\log y + ec}{l} \right)^l . \qquad \square$$

Let

$$k = \llcorner 6 \log\log y \lrcorner$$

If y is sufficiently large then

$$k > 2e \log\log y + ec,$$

and then

$$\begin{aligned} |V_k| &< \sum_{l>2k} 2^{-l} \\ &= 2^{-2k} \\ &< 2^{-12 \log\log y} \\ &< (\log y)^{-8}. \end{aligned}$$

On the other hand if we use (5.1) then we get

$$\begin{aligned} \prod_{2<p\le y} \left(1 - \frac{2}{p}\right) &= \prod_{2<p\le y} \left\{ \left(1 - \frac{1}{p}\right)^2 - \frac{1}{p^2} \right\} \\ &< \prod_{2<p\le y} \left(1 - \frac{1}{p}\right)^2 \\ &\ll \frac{1}{\log^2 y}. \end{aligned}$$

Putting all of this together we get

$$\begin{aligned} \pi_2(X) &\ll \frac{x}{\log^2 y} + \pi^{2k}(y) + \frac{x}{\log^8 y} \\ &\ll \frac{x}{\log^2 y} + \left(\frac{y}{\log y}\right)^{2k}. \end{aligned}$$

Finally, if we put

$$y = x^{1/(12 \log\log x)}$$

then by (7.1) we have

$$\begin{aligned} \pi_2(x) &\ll \frac{x(\log\log x)^2}{\log^2 x} + \left(\frac{x^{1/12 \log\log x}}{\log x / 12 \log\log x}\right)^{12 \log\log x} \\ &\ll \frac{x(\log\log x)^2}{\log^2 x}. \end{aligned}$$

□