SOME TOPICS IN PRIME NUMBER THEORY

T. N. SHOREY

Definition. An integer P is called prime if

(i) $p > 1$

(ii) p has no divisor other than 1 and itself

Examples. $2, 3, 5, 7, 13, 19, 23, \cdots$ are primes.

We denote by P the set of all primes and we write the elements of P in the increasing order as

$$p_1 < p_2 < p_3 < \cdots$$

Now consider

$$\triangle = p_1 \cdots p_n + 1.$$

Observe that $\triangle$ is an integer > 1. Therefore there exists

$$q \in P \text{ with } q \mid \triangle.$$

Then

$$q \neq p_i \text{ for } 1 \leq i \leq n$$

since $q \nmid \triangle - 1$. Thus

$$q > p_n.$$

We started with primes $p_1, \cdots, p_n$ and we found a prime $q > p_n$. Therefore we have proved

Theorem 1. (Euclid) *P is infinite.*

Let us have a closer look at the proof of Euclid. For this, we introduce the following important function for counting primes:

$$\pi(x) = \sum_{p \leq x} 1,$$

the number of primes $\leq x$. Thus

$$\pi(4) = 2, \pi(10) = 4.$$

Let $n > 1$. We have shown that

$$p_{n+1} \leq q \leq p_1 \cdots p_n + 1.$$

Therefore

$$p_{n+1} \leq p_n^n.$$

This article is based on the text of my talk in VSRP (2005) at School of Mathematics, TIFR, Mumbai.

Thus we have bounded $(n+1)$-th prime in terms of n-th prime. Using this inequality, we show that

$$p_n \leq 2^{2^{2^n}}.$$

The proof is by induction on n. It is fine when $n = 1$ and assume for n. Then

$$p_{n+1} \leq p_n^n \leq (2^{2^{2^n}})^n \leq 2^{2^{2^{n+1}}}.$$

Now

$$n = \pi(p_n) \leq \pi(2^{2^{2^n}}).$$

Let $x \geq x_0$ where x_0 is sufficiently large absolute constant. Then

$$2^{2^{2^n}} \leq x < 2^{2^{2^{n+1}}}$$

for some n. Now

$$\pi(x) \geq \pi(2^{2^{2^n}}) \geq n \geq \log\log\log x.$$

Thus Euclid's proof not only shows that primes are infinitely many in numbers but it also gives a lower bound for $\pi(x)$ and the lower bound tends to infinity with x. We have a definite result here:

Prime Number Theorem.

$$\lim_{x\to\infty} \frac{\pi(x)}{(x/\log\ x)} = 1.$$

Thus for $\epsilon > 0$,

$$\pi(x) \geq (1-\epsilon)\frac{x}{\log\ x} \text{ for } x \geq x_1(\epsilon) \tag{1}$$

and

$$\pi(x) \leq (1+\epsilon)\frac{x}{\log\ x} \text{ for } x \geq x_2(\epsilon). \tag{2}$$

Let $x = p_N$. Then

$$\lim_{N\to\infty} \frac{\pi(p_N)}{(p_N \mid \log\ p_N)} = 1$$

i.e.,

$$\lim_{N\to\infty} (\frac{N}{p_N \mid \log p_N}) = 1.$$

i.e.,

$$\lim_{N\to\infty} (\frac{p_N}{N\log N}) = 1. \tag{3}$$

Now we use Euclid's proof to show

Theorem 1′. *There are infinitely many primes of the form* $4n+3$ *with* $n > 0$.

Proof. By contradiction. Suppose that $q_1, \cdots, q_m$ are all the primes $\equiv 3 (\text{mod } 4)$. Consider

$$\triangle' = 4q_1 \cdots q_m - 1.$$

There exists $q \in P, q \mid \triangle'$ and $q \equiv 3 \pmod 4$ since $\triangle' \equiv 3 \pmod 4$. Then

$$q = q_i \text{ with } 1 \leq i \leq m$$

implying $q \mid (\triangle' + 1)$ which is a contradiction.

The above proof can not be used to show that there are infinitely many primes of the form $4n + 1$ with $n > 0$. But we have a general result.

Theorem 2. (Dirichlet). *Let $a > 0$ and $b > 0$ be integers with* $\gcd(a, b) = 1$. *Then there are infinitely many primes of the form $an + b$ with $n > 0$.*

Put

$$f(X) = aX + b.$$

Then

$$f(X) \in \mathbb{Q}[X]$$

satisfying

(i) $f(X)$ is irreducible over $\mathbb{Q}$
(ii) the leading coefficient of f is positive
(iii) Let p be a prime. Then

$$f(X) \not\equiv 0 (\text{mod } p).$$

We have the following conjecture:

Conjecture. (Schinzel) *Let $f(X) \in \mathbb{Q}[X]$ satisfying* (i), (ii) *and* (iii). *Then there are infinitely many positive integers m such that $f(m)$ is prime.*

It is easy to see that the assumptions (i), (ii) and (iii) for f are necessary. Further Schinzel formulated a more general conjecture valid for an arbitrary number of polynomials.

A powerful tool for studying primes is the Riemann Zeta function. To begin with, it is defined in the half plane

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, s = \sigma + it, \sigma > 1.$$

Its connection with primes is given by Euler Identity

$$\zeta(s) = \prod_p (1 - \frac{1}{p^s})^{-1}, s = \sigma + it, \sigma > 1.$$

$\zeta(s)$ can be continued analytically in the whole plane except at $s = 1$ where it has simple pole. We call the extended function also $\zeta(s)$. Now

$$\zeta(s) = 0 \text{ for } s = -2 - 4, \cdots$$

But these are not all the zeros of $\zeta(s)$. We have

Theorem 3. (Hardy) *There are infinitely many zeros of* $\zeta(s)$ *on* $s = \sigma + it$ *with* $\sigma = \frac{1}{2}$.

Further we have the famous conjecture:

Riemann Hypothesis. *Apart from* $s = -2, -4, \cdots$, *all the zeros of* $\zeta(s)$ *lie on the line* $s = \sigma + it$ *with* $\sigma = \frac{1}{2}$.

For $n \geq 1$, let

$$h(N) = p_{N+1} - p_n.$$

Consider the interval

$$(p_N, 2p_N].$$

It is well-known that there is $q \in P$ with $q \in (p_N, 2p_N]$. Thus

$$q \geq p_{N+1}$$

and

$$h(N) = p_{N+1} - p_N \leq q - p_N < p_N.$$

This can be improved by using Prime Number Theorem as follows: Let $0 < \theta < 1$. Then

$$h(N) < \theta N \text{ with } N \geq N_1(\theta).$$

For this we should consider the interval

$$(p_N, p_N + \theta p_N]$$

and show that it has a prime. This is equivalent to showing

$$\pi(p_N + \theta p_N) - \pi(p_N) > 0.$$

We use the lower bound (1) for the first term and upper bound (2) for the second term for deriving the above inequality. There is a

Conjecture. (Cramer) *There exists an absolute constant* N_2 *such that*

$$h(N) \leq (\log p_N)^2 \text{ for } N \geq N_2.$$

This is a very difficult conjecture. Even Riemann Hypothesis implies

$$h(N) \leq P_N^{\frac{1}{2}+\epsilon}, N \geq N_3(\epsilon).$$

It is known due to Baker, Harman and Pintz [1] that

$$h(N) \leq p_N^{\frac{1}{2}+\frac{1}{40}+\epsilon} \text{ for } N \geq N_4(\epsilon).$$

Let

$$E(N) = \frac{p_{N+1} - p_N}{\log p_N}$$

and we put

$$\begin{aligned} E_1^* &= \lim sup_{N\to\infty} E(N), \\ E_2^* &= \lim inf_{N\to\infty} E(N). \end{aligned}$$

By Prime Number Theorem

$$E_1^* \geq 1 \text{ and } E_2^* \leq 1.$$

For deriving this, we observe that

$$p_{2M} - p_M = \sum_{i=1}^{M} (p_{M+i} - p_{M+i-1})$$

and use Prime Number Theorem for getting

$$(1-\epsilon)M \log M \leq p_{2M} - p_M \leq (1+\epsilon)M \log M \text{ for } M \geq M_0(\epsilon).$$

Schönhage [9] showed that $E_1^* = \infty$ and we refer to his paper for more precise formulation of his result. There are primes of the type 3,5 or 5,7 or 11,13 These are called twin primes. It has been conjectured that there are infinitely many twin primes. Then

$$E_2^* = 0.$$

Erdős [4] was the first to show that

$$E_2^* < 1$$

and Bombieri, Friedlander, Iwaniec [2] showed that

$$E_2^* \leq \frac{6}{7}.$$

For integers $n > 0$ and $k \geq 2$, we write

$$\triangle_0 = \triangle_0(n,k) = n(n+1)\cdots(n+k-1).$$

Further we denote by $P(\triangle_0)$ and $\omega(\triangle_0)$ the greatest prime factor and the number of distinct prime divisors of $\triangle_0$, respectively. As already stated, there is a prime between X and $2X$. This is a particular case $n = k+1$ of the following result dating back to 1892.

Theorem 4. (Sylvester [12]) *We have*

$$P(\triangle_0) > k \text{ if } n > k.$$

Thus a product of k consecutive positive integers each exceeding k is divisible by a prime greater than k. By applying Theorem 4 with $n = k+1$, we have

$$P(\triangle_0(k+1,k)) > k.$$

Therefore there is an integer between $k+1$ and $2k$ divisible by a prime exceeding k and this integer has to be a prime. The assumption $n > k$ in Theorem 4 is necessary since

$$P(\triangle_0(1,k)) = P(1.2\cdots k) \leq k.$$

We have several more such instances. Let $n > 1$ and $k = n! + 1$. We write

$$\triangle_0(n,k) = n(n+1)\cdots(n!+1)(n!+2)\cdots(n!+n)$$

and we observe that

$$P(\triangle_0) \leq n! + 1 = k,$$

since $n! + 2, \cdots, n! + n$ are all composites. Thus there are infinitely many pairs (n, k) for which

$$P(\triangle(n, k)) \leq k.$$

This is special about consecutive integers. Let $d > 1$, gcd $(n, d) = 1$ and $k \geq 3$. Then Shorey and Tijdeman [11] showed that

$$P(n(n + d) \cdots (n + (k - 1)d)) > k$$

unless $(n, d, k) = (2, 7, 3)$. We observe that $P(2.9.16) = 3$ and therefore it is necessary to exclude the tuple (2,7,3). Also the assumption $k \geq 3$ is necessary since

$$P(1(1 + 2^r - 1)) = P(1.2^r) = 2 \text{ for } r = 1, 2, \cdots .$$

We know that

$$k! \mid \triangle_0(n, k).$$

Thus

$$\omega(\triangle_0(n, k)) \geq \pi(k).$$

Theorem 4 can be re-formulated as

$$\omega(\triangle_0) > \pi(k) \text{ if } n > k. \tag{4}$$

Let us see how far we can go. We observe that

$$\triangle_0(k + 1, k) = (k + 1) \cdots (2k)$$

and

$$\omega(\triangle_0(k + 1, k)) = \pi(k) + \pi(2k) - \pi(k) = \pi(2k)$$

In fact, we can say a little more. We consider

$$\triangle_0(k + 2, k) = (k + 1)(k + 2) \cdots (2k)\frac{(2k + 1)}{(k + 1)}$$

and

$$\omega(\triangle_0(k + 2, k)) = \pi(k) + \pi(2k) - \pi(k) - 1 = \pi(2k) - 1$$

if $k + 1$ is prime and $2k + 1$ is composite. There are infinitely many such k. We have already seen that there are infinitely many primes $p \equiv 2 (\text{mod } 3)$. Let $k = p - 1$. Then $k + 1 = p$ is prime and

$$2k + 1 = 2(k + 1) - 1 \equiv 0 (\text{mod } 3)$$

implying the assertion. There are examples when $\omega(\triangle_0(n, k)) < \pi(2k) - 1$. For example

$$\omega(\triangle_0(74, 57)) = \pi(2k) - 2, \omega(\triangle_0(3936, 3879)) = \pi(2k) - 3,$$
$$\omega(\triangle_0(1304, 1239)) = \pi(2k) - 4, \omega(\triangle_0(3932, 3880)) = \pi(2k) - 5$$

but we do not know whether there are infinitely many such pairs. Sylvester's theorem can not be sharpened to

$$\omega(\triangle_0(n, k)) \geq \pi(2k) \text{ for } n > k.$$

On the other hand, Laishram and Shorey [6] showed that

$$\omega(\triangle_0(n,k)) \geq \min(\pi(k) + [\frac{3}{4}\pi(k)] - 1, \pi(2k) - 1) \text{ for } n > k.$$

If the interval $[n, n+k)$ is contained in an interval (p_N, p_{N+1}), then the estimate (4) may be sharpened considerably. Infact Grimm [5] conjectured that

$$\omega(\triangle_0) \geq k$$

if $n, n+1, \cdots, n+k-1$ are all composites. This conjecture, according to Erdős, implies that there exist absolute constants $\alpha > 0$ and N_5 such that

$$p_{N+1} - p_N < p_N^{\frac{1}{2}-\alpha} \text{ for } N \geq N_5.$$

The conjecture has been confirmed by Ramachandra, Shorey and Tijdeman [8] when $\log k \leq C_1(\log\ n)^{1/2}$ for $n \geq N_6$ where C_1 and N_6 are absolute constants.

Now we point out a relation between Theorem 4 and Diophantine equations. Let k be fixed and $P(\triangle_0) \leq k$. Then the terms of $\triangle_0$ are composed of fixed primes. For any three distinct terms $n+i_1, n+i_2$ and $n+i_3$ of $\triangle_0$ with $i_1 < i_2 < i_3$, we have Siegel's identity

$$(i_3 - i_2)(n+i_1) + (i_2 - i_1)(n+i_3) = (i_3 - i_1)(n+i_2).$$

Therefore we have an equation of the form

$$X_1 + X_2 = X_3$$

where X_1, X_2 and X_3 are composed of primes from a given set. This leads us to a fundamental and central problem in Diophantine equations:

a b c Conjecture. Let a, b *and* c *be positive integers such that* $\gcd(a,b,c) = 1$ *and*

$$a + b = c.$$

Let $\epsilon > 0$. *Then there exists* $K = K(\epsilon)$ *such that*

$$c \leq K(\prod p)^{1+\epsilon}$$

where the product is taken over all prime divisors of abc.

This conjecture has several consequences. For example, it can be applied to Fermat equation (5) to obtain the following result.

Theorem 5. *Let* $p \geq 3$ *be prime and* x, y, z *be positive integers such that* $\gcd(x,y,z) = 1$ *and*

$$x^p + y^p = z^p. \tag{5}$$

Then

$$\max(p, x, y, z) \leq C_2$$

where C_2 *is an absolute constant.*

Now we show that *a b c* conjecture implies Theorem 5.

Proof. Assume (5). Then $p > 3$ by Euler. We apply $a\ b\ c$ conjecture with

$$a = x^p, b = y^p, c = z^p, \epsilon = \frac{1}{6}.$$

Then gcd $(a, b, c) = 1$ and $a + b = c$. We obtain

$$z^p \leq K(\prod_{p|xyz} p)^{7/6} \leq K(xyz)^{7/6} \leq Kz^{7/2}$$

where K is an absolute constant. Thus $2^{p-\frac{7}{2}} \leq z^{p-\frac{7}{2}} \leq K$. Hence p and z are bounded since $p > 3$. Consequently x and y are bounded.

Wiles [13] has proved that Fermat's equation does not hold. Now we state a result proved recently coming out of the ideas of Wiles and others on Fermat's equation and the theory of linear forms in logarithms. We define the Fibonacci sequence

$$F_0 = 0, F_1 = 1$$

and

$$F_n = F_{n-1} + F_{n-2} \text{ for } n \geq 2.$$

We write the members of the sequence

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \cdots .$$

Thus $F_0 = 0, F_1 = 1, F_2 = 1, F_6 = 8, F_{12} = 144$ are powers. It has been proved recently by Bugeaud, Mignotte and Siksek [3] that these are the only powers in the Fibonacci sequence.

Fibonacci sequences are binary recursive sequences. Let $u_0, u_1 \in \mathbb{Q}$ and $r, s \in \mathbb{Q}$ with $s \neq 0, r^2 + 4s \neq 0$. Then we consider the binary recursive sequence $\{u_m\}$ given by

$$u_m = ru_{m-1} + su_{m-2} \text{ for } m \geq 2.$$

Let α and β be roots of $X^2 - rX - s$. Then $\alpha\beta \neq 0, \alpha \neq \beta$ and $\alpha + \beta = r, \alpha\beta = -s$. Now we show by induction on m that

$$u_m = a\alpha^m + b\beta^m \text{ for } m \geq 0$$

where

$$a = \frac{u_0\beta - u_1}{\beta - \alpha}, b = \frac{u_1 - u_0\alpha}{\beta - \alpha}.$$

Then $\{u_m\}$ is called non-degenerate if $ab \neq 0$ and α/β is not a root of unity. It has been proved by Pethő [7] and Shorey and Stewart [10], independently, that there are only finitely many powers in a non-degenerate binary recursive sequence and the proof depends on the theory of linear forms in logarithms.

References

[1] R.C. Baker, G. Harman and J. Pintz, *The difference between consecutive primes* II, Proc. London Math. Soc. 83 (2001), 532-562.

[2] E. Bombieri, J.B. Friedlander and H. Iwaniec, *Primes in arithmetic progressions to large moduli,* Acta Math. 156 (1986) , 203-251.

[3] Y. Bugeaud, M. Mignotte and S. Siksek, *Classical and modular approaches to exponential diophantine equations* I, Fibonacci and Lucas perfect powers, Annals of Math., to appear.
[4] P. Erdős, *The difference of consecutive primes*, Duke Math. Journal 6 (1940), 438-441.
[5] C.A. Grimm, *A conjecture on consecutive composite numbers*, American Math. Monthly 76 (1969), 1126-1128.
[6] Shanta Laishram and T.N. Shorey, *Number of prime divisors in a product of consecutive integers*, Acta Arith. 113 (2004), 327-341.
[7] A. Pethő, *Perfect powers in second order linear recurrences,* Journal Number Theory 15 (1982), 5-13.
[8] K. Ramachandra, T.N. Shorey and R. Tijdeman, *On Grimm's problem relating to factorisation of a block of consecutive integers* II, Journal Reine Angew Math. 288 (1976), 192-201.
[9] Arnold Schönhage, *Eine Bemerkung zur Konstruktion grosser Primzahllücken*, Arch. Math. 14 (1963), 29-30.
[10] T.N. Shorey and C.L. Stewart, *On the equation* $ax^2 + bx^t y + cy^2 = d$ *and pure powers in recurrence sequences*, Math. Scand. 52 (1983), 24-36.
[11] T.N. Shorey and R. Tijdeman, *On the greatest prime factor of an arithmetical progression*, A tribute to Paul Erdős, ed. A. Baker, B. Bollobas and A. Hajnal, Cambridge University Press (1990), 385-389.
[12] J. Sylvester, *On arithmetical series*, Messenger Math. 21 (1892), 1-19, 87-120.
[13] A. Wiles, *Modular elliptic curves and Fermat's Last Theorem* , Annals of Math. 141 (1995), 443-551.

School of Mathematics, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India

E-mail address: shorey@math.tifr.res.in