PRIMES AND TWIN PRIMES

Vesta Coufal

1992

How many primes are there? How are they distributed among the integers? How can one tell if a given integer is a prime? Is there a formula which produces all primes, or only primes? Primes greater then two are all odd (since two divides all even integers), so what about pairs (p, p+2) where p and p+2 are both prime (these are called twin primes)? How many of these pairs are there? How are they distributed in the integers and in the primes?... Many such questions have been asked and many conjectures made about the sequence of prime numbers, but even now, after the hundreds of years these numbers have been studied, relatively few significant results have been proved. Actually, now there are often a number of different proofs for the same result. Unfortunately, in most cases the ideas in the proofs don't lead anywhere else of significance. Furthermore as the theory progresses, the questions and answers remain fairly basic while the proofs quickly leave the realms of elementary mathematics.

In this paper, I will present some of the answers and partial answers - the theory - surrounding the primes and twin primes, and my observations and data about the twin primes. First, though, some notes on notation are needed:

- $\star$ p and p_n will always denote prime numbers with p_n being the n_{th} prime and $p_1 = 2$.
- $\star$ $\prod_p f(p)$ and $\sum_p f(p)$ are, respectively, the product and sum of $f(p)$ over all the primes.
- $\star$ $\sum_{j|n}$ is the sum on j where j runs through all the positive divisors of a positive integer n.
- $\star$ $f \sim g$ is read "f is asymptotic to g" and means that $\lim_{x\to\infty} \frac{f}{g} = 1$.
- $\star$ $\lfloor x \rfloor$ is the integral part of x (that is, the integer n for which $n \leq x < n+1$).
- $\star$ For non-negative functions f and g, $f(x) = O(g(x))$ if there exists a $K > 0$ such that $f(x) < Kg(x)$ for large x.

PRIMES

One of the first major results concerning prime numbers is:

Theorem 1 (Euclid's Second Theorem). *The number of primes is infinite.*

This was proved by Euclid as follows:

(proof of Theorem 1)

Let 2,3,5,...,p_n be all the primes up to p_n, and let q_n be the sum of one and the product of all these primes ($q_n = 2 \cdot 3 \cdot 5 \cdots p_n + 1$). It is apparent that q_n is not divisible by any of 2,3,5,...,p_n since none of these divides one. Therefore, q_n must be prime or divisible by a prime between p_n and q_n. This shows that there is a prime greater than p_n for any $n > 0$, and thus that there are infinitely many primes. $\diamond$

Euclid's proof is an example of a proof that doesn't seem to lead anywhere useful other than the theorem itself. For example, from the construction of q_n above, $p_m | q_n$ for some $p_{n+1} \leq p_m \leq q_n$. It follows, by induction, that $p_n < 2^{2^n}$. However, as an upper bound on the n^{th} prime, this is not very effective. For instance, this says that $p_2 < 16$, but $p_2 = 3$. As n increases, the upper bound gets even worse ($p_8 = 19$ while the upper bound is given as about 1.16×10^{77})!

There are suprises, though. Euler later approached the Theorem by a new method, by showing:

Theorem 2. *The series $\sum_p \frac{1}{p}$ diverges.*

Before I can prove Theorem 2, I need:

Theorem3. *The product $\prod_p \frac{1}{1-\frac{1}{p}}$ over all primes p diverges.*

(Proof of Theorem 3)

For large N,

$$\begin{aligned}
\prod_{p \leq N} \frac{1}{1-\frac{1}{p}} &= \prod_{p \leq N} (1 + \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \cdots) \\
&= \prod_{p \leq N} \sum_{i=0}^{\infty} \frac{1}{p^i} \\
&\geq \prod_{p \leq N} \sum_{p^i \leq N} \frac{1}{p^i} \\
&\geq \sum_{n=1}^{N} \frac{1}{n} \qquad (1) \\
&\geq \int_1^N \frac{1}{x} \\
&> \log N
\end{aligned}$$

Thus, $\lim_{x \to \infty} \log N$ diverges, implying Theorem 3. $\diamond$

Now for Theorem 2.

(Proof of Theorem 2)

Taking the log of (1) gives:

$$\log \sum_{n=1}^{N} \frac{1}{n} \leq \log \prod_{p \leq N} \frac{1}{1 - \frac{1}{p}}$$

$$= - \sum_{p \leq N} \log \left(1 - \frac{1}{p}\right)$$

$$\leq - \sum_{p \leq N} \log(1 - \frac{1}{p})$$

$$= \sum_{p \leq N} \frac{1}{p} + R(N)$$

where $R(N) = \sum_{p \leq N} \sum_{m=2}^{\infty} \frac{1}{mp^m}$ and $\log(1 - \frac{1}{p})$ is replaced by its Taylor expansion. Now consider $R(N)$. $R(N) > 0$ and

$$R(N) < \sum_{p \leq N} \sum_{m=2}^{\infty} \frac{1}{p^m}$$

$$= \sum_{p \leq N} \frac{1}{p^2 - p}$$

$$< 2 \sum_{p \leq N} \frac{1}{p^2}$$

$$< 2 \sum_{n \leq N} \frac{1}{n^2}$$

$$< 2 \sum_{n=1}^{\infty} \frac{1}{n^2}$$

Thus, since $2 \sum_{n=1}^{\infty} \frac{1}{n^2}$ is bounded, $R(N)$ is also bounded by L (say). It follows that

$$\log \sum_{n=1}^{\infty} \frac{1}{n} - L < \sum_{p} \frac{1}{p}$$

Since the left side diverges as x increases, so does $\sum_p \frac{1}{p}$. $\diamond$

This also proves Theorem 1.

The supprise here is (1). This follows from an identity of Euler's:

$$\sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p} (1 + p^{-s} + p^{-2s} + \cdots) v = \prod_{p} \frac{1}{1 - \frac{1}{p^{-s}}}$$

From this identity comes:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$$

called the Riemann zeta-function because of his extensive contributions to the understanding of the function. It was using this function and the knowledge surrounding it that the next, and perhaps most important result about primes was first proved.

Theorem 4 (The Prime Number Theorem).

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

where $\pi(x)$ is the number of primes less than or equal to x.

However, Theorem 4 also has other proofs that do not use the Reimann zeta-function. One of them is presented in "A Motivated Account of an Elementary Proof of the Prime Number Theorem" by Norman Levinson [4]. Here, 'elementary' means the proof avoids using complex variables - it is far from simple. The groundwork and the proof itself extends over 20 pages of interwoven formulae. Obviously, a full account of the paper is not appropriate here, but a discussion of the general ideas should be interesting.

Levinson begins the paper by laying the groundwork for the actual proof. In the first section he introduces a number of functions and inequalities related to prime numbers. He starts with the Fundamental Theorem of Arithmetic which states that all positive integers can be factored uniquely into a product of powers of primes. That is, $n = p_1^{r_1} + p_2^{r_2} + \cdots + p_m^{r_m}$, where the p_j are distinct primes and the r_j are positive integers. It follows that

$$\log n = \sum_{j|n} \Lambda(j)$$

where

$$\Lambda(n) = \log p$$

for $n = p^i$ and $\Lambda(n) = 0$ otherwise. The weight $\log p$ (and thus Λ and related functions) is a good tool for finding the asymptotic behavior of the primes and it stems from the observation by Gaus that the density of the primes at x is about $\frac{1}{\log x}$ Further,

$$\psi(x) = \sum_{j \le x} \Lambda(j) = \sum_{p^m \le x} \log p \tag{2}$$

is a weight which is obviously closely related to $\Lambda(x)$ and gives the Chebychef identity

$$T(x) = \sum_{ij \leq x} \Lambda(j) = \sum_{n \leq x} \psi(\frac{x}{n}). \tag{3}$$

In studying the behavior of these fundtions, and thus the primes, it will be necessary to have expressions for Λ and ψ in terms of $\log(x)$ and $T(x)$ respectively. For this purpose, Levinson introduces the Möbius inversion formula, which is defined as follows: for functions $F(x)$ and $G(x) = \sum_{n \leq x} F(\frac{x}{n})$, both defined for $x > 1$, the inversion formula is $F(x) = \sum_{k \leq x} \mu(k) G(\frac{x}{k})$, where $\mu(n) = (-1)^m$ if $n = p_1 p_2 \cdots p_m$ for distinct primes p_j, and $\mu(n) = 0$ if $p^2 | n$. Thus, applying the Möbius inversion formula to (3) gives

$$\psi(x) = \sum_{k \leq x} \mu(k) T(\frac{x}{k}), \tag{4}$$

and it follows from the definitions of $\psi(x)$ and $T(x)$ that

$$\Lambda(n) = \sum_{k|n} \mu(k) \log(\frac{n}{k}) \tag{5}$$

for $n \geq 1$.

In the second section Levinson proves a number of important elementary results.

Lemma 1. *Let $f(t)$ have a continuous derivative, $f'(t)$, for $t \geq 1$. Let $c_n, n \geq 1$, be constants and let $C(u) = \sum_{n \leq u} c_n$. Then*

$$\sum_{n \leq x} c_n f(n) = f(x) C(x) - \int_1^x f'(t) C(t)\, dt$$

and

$$\sum_{n \leq x} f(n) = \int_1^x f(t)\, dt + \int_1^x (t - \lfloor x \rfloor) f'(t)\, dt + f(1) - (x - \lfloor x \rfloor) f(x). \tag{6}$$

For $f(t) = \log t$, (6) gives

$$T(x) = x \log x - x + O(\log x) \tag{7}$$

which is an alternate, and often more informative expression for $T(x)$ than (3).

Using (2), (3), a change of variable and a bit of arguing gives the following upper bound on $\psi(x)$.

Lemma 2. *For large x*

$$\psi(x) < \frac{3}{2} x.$$

This upper bound is useful in proving the next three lemmas and is used occasionally throughout the rest of the paper.

Lemma 3.

$$\sum_{n\leq x}\frac{\Lambda(n)}{n}=\log x+O(1).$$

This lemma is necessary only for proving that a crucial expression later on is bounded (Lemma 12).

Lemma 4.

$$\psi(x)=\pi(x)\log x+O(\frac{x\log(\log x)}{\log x}),$$

so that $\lim_{x\to\infty}\frac{\psi(x)}{x}=1$ *is equivalent to the Prime Number Theorem.*

Lemma 4 is the first of a number of restatements of the goal - proving that $\lim_{x\to\infty}\frac{\pi(x)}{\frac{x}{\log x}}=1$. Thus, the goal is now to prove that $\lim_{x\to\infty}\frac{\psi(x)}{x}=1$. Finally, Lemma 5 is used only in developing the results of the third section, and it comes from applying (6) to $f(t)=\frac{1}{t}$. Note that $0<\gamma<1$.

Lemma 5.

$$\sum_{n\leq x}\frac{1}{n}=\log x+\gamma+O(\frac{1}{x}),$$

where γ *is Euler's constant*

$$\gamma=1-\int_1^\infty\frac{t-\lfloor t\rfloor}{t^2}\,dt.$$

Continuing, in the third section he uses (4), (7), and Lemma 5 to derive the crude result

$$\psi(x)=O(x).$$

Actually, this follows directly from Lemma 2, but the method used to get it here, along with the crude result itself, (4), Lemma 3, some substitutions and a few other pieces of information, also produces Selberg's inequality

$$\psi(x)\log x+\sum_{n\leq x}\Lambda(n)\psi(\frac{x}{n})=2x\log x+O(x).$$

Using various previous results, Selberg's inequality, and setting $\Lambda_2(n)=\Lambda(n)\log n+\sum_{jk=n}\Lambda(j)\Lambda(k)$ gives

$$Q(n)=\sum_{k\leq n}(\Lambda_2(k)-2\log k)=O(n),$$

for $n\geq 2$, and $Q(1)=0$, an equation which is used only for Lemma 8.

Finally, in the fourth section Levinson presents the actual proof of the Prime Number Theorem. He begins by defining the function $R(x) = \psi(x) - x$, $x \geq 2$, and $R(x) = 0$, $x < 2$, for which

$$R(x)\log x + \sum_{n\leq x}\Lambda(n)R(\frac{x}{n}) = O(x). \tag{8}$$

From this, the goal is again redefined to be

$$\lim_{x\to\infty}\frac{R(x)}{x} = 0.$$

Next, in order "to get a more tractable inequality," he performs a series of smoothing operations on (3). First replace $R(x)$ by

$$S(y) = \int_2^y \frac{R(x)}{x}\,dx \tag{9}$$

if $y \geq 2$ and $S(y) = 0$, $y < 2$, which is a smoother function. Some relevant information about (9) is given in

Lemma 6. *There exists a constant c such that*

$$|S(y)| \leq cy$$

for $y \geq 2$, and

$$|S(y_2) - S(y_1)| \leq c|y_1 - y_2|.$$

Moreover a consequence of (3) is

$$S(y)\log y + \sum_{j\leq y}\Lambda(j)S(\frac{y}{j}) = O(y).$$

These inequalities are proved using lemma 2, (8), (9), and integration.

Lemma 7. *With $\Lambda_2(n)$ defined as before and K_1 a constant*

$$\log^2 y|S(y)| \leq \sum_{m\leq y}\Lambda_2(m)|S(y)| + K_1 y\log y.$$

Lemma 8. *There is a constant K_2 such that*

$$\log^2 y|S(y)| \leq 2\sum_{m\leq y}|S(\frac{y}{m})|\log m + K_2 y\log y.$$

Lemma 9. *There is a constant K_3 such that*

$$\log^2 y|S(y)| \leq 2\int_x^y |S(\frac{y}{u})| \log u \, du + K_3 y \log y.$$

The previous three lemmas perform further smoothing operations, and, with a little more work and letting $W(x) = e^{-x}S(e^x)$, they culminate in

$$|W(x)| \leq \frac{2}{x^2}\int_0^x (x-v)|W(v)|\, dv + \frac{K_3}{x},$$

and the following important lemma:

Lemma 10. *Let*

$$\alpha = \limsup_{x\to\infty} |W(x)|,$$

$$\gamma = \limsup_{x\to\infty} \frac{1}{x}\int_0^x |W(\xi)|\, d\xi;$$

then $\alpha \leq 1$ and

$$\alpha \leq \gamma.$$

This further redefines the goal to be $\alpha = 0$. However, before this can be shown, two more crucial facts, Lemmas 11 and 12, are needed about $W(x)$.

Lemma 11. *If $k = 2c$ then*

$$|W(x_2) - W(x_1)| \leq k|x_2 - x_1|,$$

and hence

$$||W(x_2)| - |W(x_1)|| \leq k|x_2 - x_1|.$$

This follows directly from Lemma 6 and its proof.

Lemma 12. *If $W(v) \neq 0$ for $v_1 < v < v_2$, then there exists a number M such that*

$$\int_{v_1}^{v_2} |W(v)|\, dv \leq M,$$

where $W(v) \neq 0$ and $v_1 < v < v_2$.

Lemmas 1 and 2, a change of variable and various manipulations prove this lemma. Finally, there is enough information to prove

Lemma 13. *A function $W(x)$ subject to the three conditions in Lemmas 10, 11, and 12 must in fact have $\alpha = 0$.*

Thus, working back through the progression of restatements of the goal, the Prime Number Theorem is proved!

Since the above is only an outline of the proof, it is easy to see that the proof itself is quite involved, even though the statement of the theorem is so simple. Further results about primes are messier to state and their proofs involve much deeper results of number theory.

TWIN PRIMES

Twin primes (pairs $(p, p+2)$ such that both p and $p+2$ are prime) are, in a sense, even more obscure than primes. Consequently, there are fewer results concerning twin primes. In fact, very little is known about this sequence of numbers. Although data and heuristic arguments seem to support it, the following basic idea is still an outstanding conjecture:

Conjecture 1. *The number of twin primes is infinite.*

As with the primes, let $\pi_2(x)$ be the number of twin primes $(p, p+2)$ such that $p \leq x$. If the conjecture is correct, $\pi_2(x)$ should tend to infinity as x tends to infinity. This is supported by comparisons of x and $\pi_2(x)$ (see Table 1, page 12), the fact that larger and larger twin primes are still being found (the largest pair as of 1985 is $107570463 \times 10^{2250} \pm 1$, which has 2259 digits! [5]), and the following conjecture made by Hardy and Littlewood [2]:

Conjecture 2.

$$\pi_2(x) \sim \frac{2C_2 x}{\log^2 x}$$

where $C_2 = \prod_{p \geq 3} (1 - \frac{1}{(p-1)^2})$.

(C_2 is called the twin prime constant. $C_2 \approx 0.66016\ldots$ [5]) For a comparison of $\pi_2(x)$ and results of Conjecture 2, see Table 1, page 12. Furthermore, there is a heuristic argument for Conjecture 2 which, if correct, would prove Conjecture 1. However, this argument has not yet been tightened into a proof.

Recalling the primes, where one result which proved the infinitude of primes was that $\sum_p \frac{1}{p}$ diverges, it would seem reasonable to try the same thing for the twin primes. Unfortunately, this approach is rendered useless by

Theorem 5.

$$\sum (\frac{1}{p} + \frac{1}{p+2})$$

summed over twin prime pairs $(p, p+2)$ *converges.*

(For the proof of Theorem 5 see [5] page 401)

This result was proved by Run in 1919. Thus, $B = \sum(\frac{1}{p} + \frac{1}{p+2}) = (\frac{1}{3} + \frac{1}{5}) + (\frac{1}{5} + \frac{1}{7}) + (\frac{1}{11} + \frac{1}{13}) + \cdots =$ 1.90216054... is called Brun's constant[5]. Theorem 5 indicates that even if there are an infinitude of twin primes, they are relatively sparse.

The above conjectures and Theorem 5 are essentially all I could find in the literature about twin primes. However, many of my own questions concerned how the twin primes were distributed among the primes. What do they look like as they thin out (as they increase in size)? Is there any patterns? Or are they distributed randomly among the primes? These questions led me to collect some data of my own.

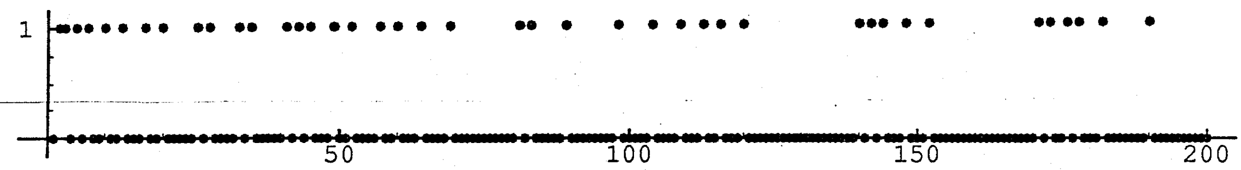
The twin graphs (see pages 13, 14, and 15 for a sampling of the graphs) plot the position of twin primes in relation to the primes. The primes p_n are the points on the x-axis (there are 200 such points on each graph) and there is a point at 1 for each pair of twin primes. I could find no clear pattern in the graphs, so I constructed similar graphs of randomly generated numbers to see how they compared to the twin graphs. To do this, I generated random numbers $r_1, r_2, r_3, \ldots$ with the same density as the primes $(\frac{1}{\log x})$. Next, I graphed the random "twins" (pairs of random numbers $(r, r+2)$ such that r and $r+2$ are prime) in the same way as the twin primes.

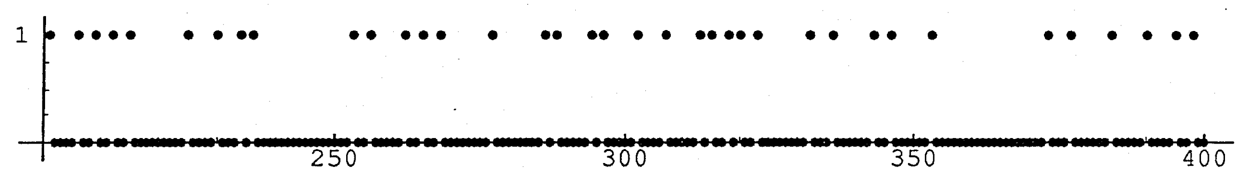
The first problem I encountered was that the only occurrence of successive twin primes $(p, p+2), (p+2, p+4)$ is for $p = 3$. (To see that this is true, consider $p > 3$. Then $p \equiv 1(\text{mod}4)$ or $p \equiv 3(\text{mod}4)$. But for each of these cases, $p+2$ and $p+4$ cannot both be prime.) However, for the random twins, this occurred relatively often and removing them lowered their density noticably. However, it was already fairly apparent that, excluding the density problems, there was no discernible difference between the distribution of the twin primes and the random twins.

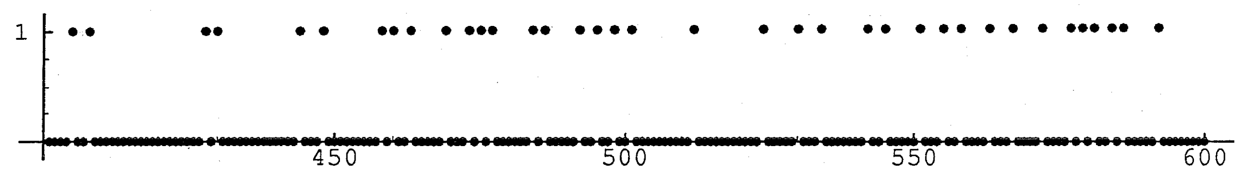
An incredible amount of work has been done concerning the prime numbers, but relatively few significant result have been proved. Concerning the twin primes, less work has been done and proportionately few results exist. Thus, further study of these sequences of numbers and the theory surrounding them would certainly be interesting and useful.

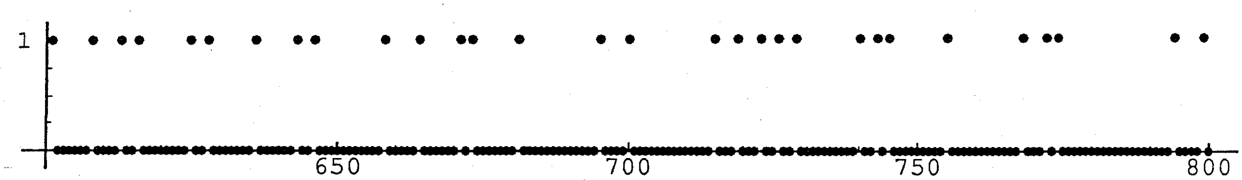
TABLE 1

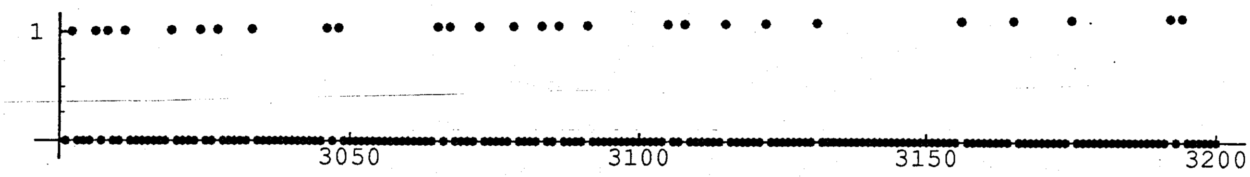
x	$\pi_2(x)$	$\frac{2C_2 x}{\log^2 x}$	$\frac{\pi_2(x)}{\frac{2C_2 x}{\log^2 x}}$
10^3	35	28	1.25
10^4	205	156	1.3141
10^5	1224	996	1.22892
10^6	8169	6917	1.181
10^7	58980	50822	1.16052
10^8	440312	389107	1.1316
10^9	3424506	3074426	1.11387
10^{10}	27412679	24902847	1.10078
10^{11}	224376048	205808657	1.109021

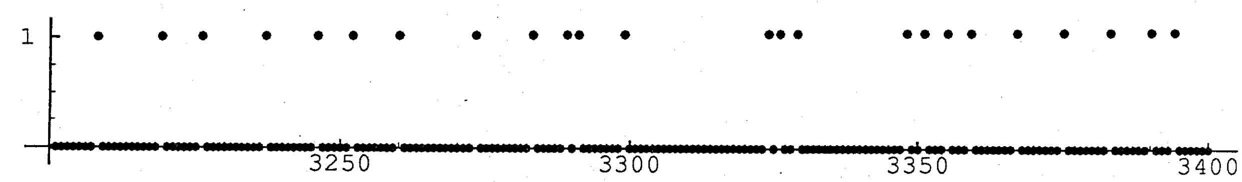

1
50
100
150
200

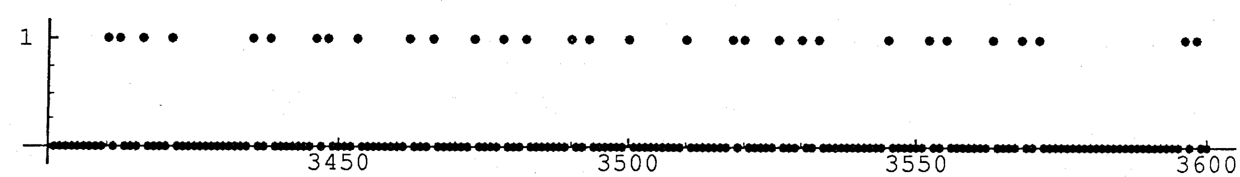

1
250
300
350
400

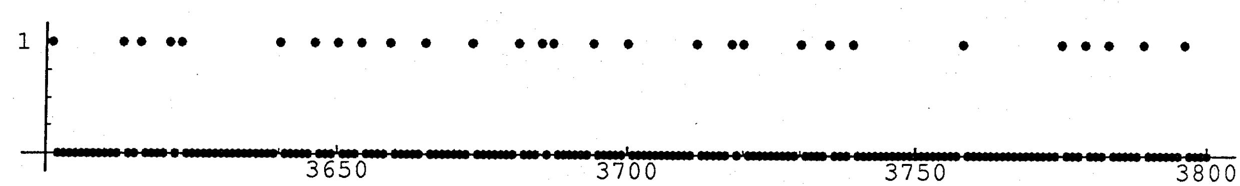

1
450
500
550
600

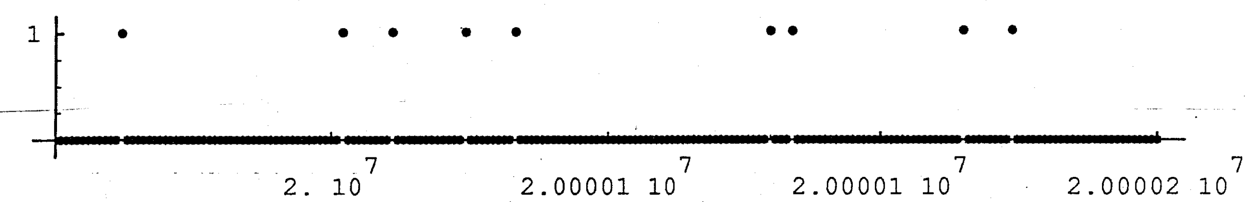

1
650
700
750
800

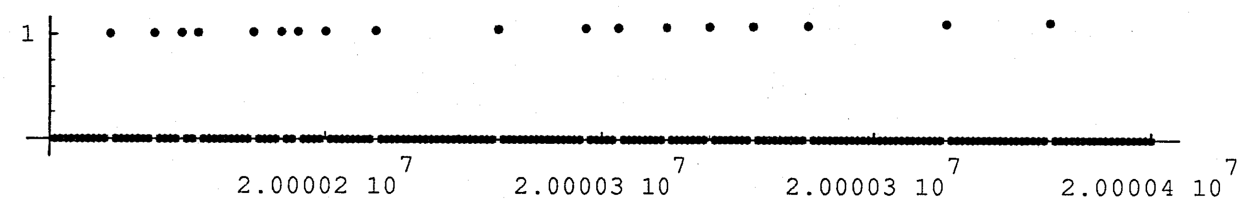

1
3050
3100
3150
3200

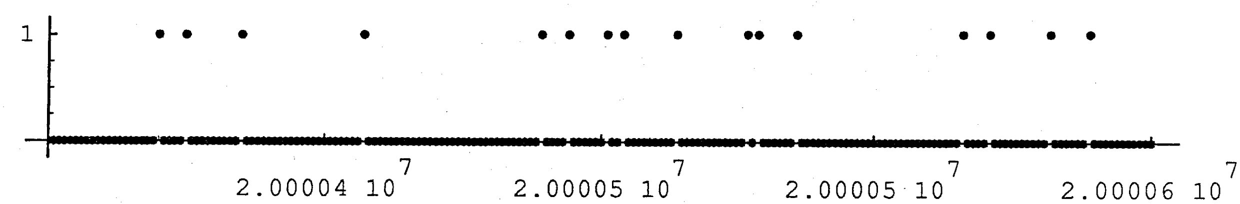

1
3250
3300
3350
3400

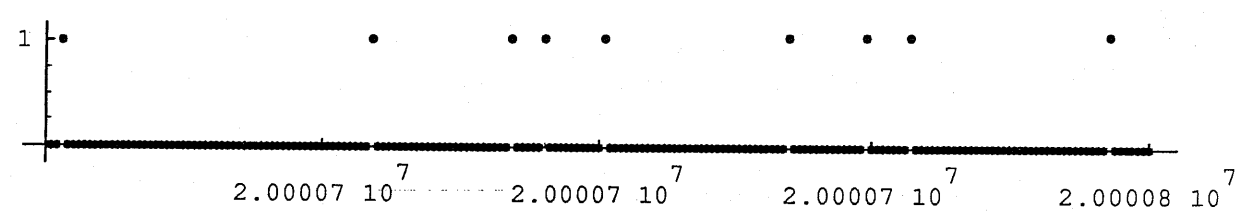

1
3450
3500
3550
3600


1
3650
3700
3750
3800


1
2. 10^7
2.00001 10^7
2.00001 10^7
2.00002 10^7


1
2.00002 10^7
2.00003 10^7
2.00003 10^7
2.00004 10^7


1
2.00004 10^7
2.00005 10^7
2.00005 10^7
2.00006 10^7


1
2.00007 10^7
2.00007 10^7
2.00007 10^7
2.00008 10^7

REFERENCES

1. Emil Grosswald, Topics from the Theory of Numbers, Birkhäuser, Boston, 1984.
2. G. H. Hardy and E. M. Wright, An Introduction to the Theory of Numbers, Oxford University Press, New York, 1960.
3. A. E. Ingham, The Distribution of Prime Numbers, Cambridge University Press, London, 1932.
4. Norman Levinson, A Motivated Account of an Elementary Proof of the Prime Number Theorem, American Math Monthly, 76 (1969) 225-245.