Separatum
ELEMENTE DER MATHEMATIK
BIRKHÄUSER VERLAG BASEL, SCHWEIZ
Vol. 35/5, 1980

Remarks on the differences between consecutive primes

In problem 654, Journal of Recreational Mathematics, Harry Nelson asks: "What is the most likely difference between consecutive primes?" Here a difference is 'most likely' for primes $\leq n$ if it occurs at least as often as any other difference. For a discussion see J. Rec. Math. *11*, 231.

We first show that a well-known conjecture of Hardy and Littlewood implies that the most likely difference tends to infinity with n, so that there is no most probable difference independent of n.

In Hardy and Littlewood: On the expression of a number as a sum of primes. Collected Papers of G.H. Hardy, VI, p. 682; they conjecture that the number of solutions of

$$p_i - p_j = 2k, \qquad p_i \le n, \tag{1}$$

equals

$$\left(c + o(1)\right) \frac{n}{\log^2 n} \prod_{\substack{p \mid k \\ p \text{ odd}}} \frac{p-1}{p-2} \tag{2}$$

where c is an absolute constant.
Let p_i denote the ith prime; then (2) implies that the number of solutions of

$$p_{i+1} - p_i = 2k, \qquad p_{i+1} \le n \tag{3}$$

also is of the form (2). Since every solution of (3) is a solution of (1) it is clear that the number cannot be greater than (2). On the other hand if we have a solution $p_i - p_j = 2k$ of (1) which is not a solution of (3), that is $i > j+1$, then we get a triple of primes

$$p_j, \qquad p_j + 2u, \qquad p_j + 2k; \qquad 1 \le u < k. \tag{4}$$

From Brun's method it follows that the number of such triples with $p_j < n$ is less than

$$c_k n \prod_{k < p < n^{\varepsilon}} \left(1 - \frac{3}{p}\right) < c_k' \frac{n}{\log^3 n} \qquad (\varepsilon > 0) \tag{5}$$

for each fixed u, and hence $\le c_k'' n / \log^3 n$ for all triples in (4). Inequality (5) follows from the fact that the primes satisfying (4) exclude three residue classes $(\bmod\, p)$ for $p > k$. Since the bound (5) is small compared to the estimate (2) it follows that (2) is also an estimate for the number of solutions in (3).
Now (2) implies that the most likely difference between consecutive primes goes to infinity with n. Denote the number of solutions of (3) by $f(n,k)$ and let k_n be the minimum value of k for which $f(n,k)$ is maximal.
Brun's method gives the well-known relation

$$f(n,k) < c_k n / \log^2 n \tag{6}$$

In view of the divergence of $\prod (p-1)/(p-2)$ estimates (2) and (6) imply that

$$f(n, k_n) \Big/ \frac{n}{\log^2 n} \to \infty \tag{7}$$

and (6), (7) imply $k_n \to \infty$ with n.

Of course the prime number theorem implies

$$f(n,k_n)\bigg/\frac{n}{\log^2 n}>c>0 \tag{8}$$

for some fixed constant c, but this is not sufficient to prove that $k_n\to\infty$.
Next we ask: How fast does k_n go to infinity? We conjecture that

$$k_n/(\log n)^{1-\varepsilon}\to\infty \quad \text{for every} \quad \varepsilon>0, \tag{9}$$

but

$$k_n/\log n\to 0. \tag{10}$$

Conjecture (2) is not strong enough to deduce (9) or (10). Perhaps they can be deduced from stronger plausible conjectures.
Let l_n be the largest integer for which

$$f(n,k_n)=f(n,l_n) \tag{11}$$

then we still expect that

$$l_n/\log n\to 0. \tag{12}$$

Finally we conjecture that

$$f(n,k_n)\log^2 n/(n\log\log n)\to c>0. \tag{13}$$

Without unproven conjectures we cannot even improve (8) to

$$f(n,k_n)\log^2 n/n\to\infty. \qquad (14$$

Let $F(n,k)$ denote the number of solutions of

$$p_i-p_j=2k, \qquad p_i\le n. \tag{15}$$

Let K_n be the least integer k for which $F(n,k)$ is maximal.
Then

$$F(n,K_n)>c\,\frac{n\log\log n}{\log^2 n}\,. \tag{16}$$

In order to prove (16) choose $A=p_1\cdots p_m\le\sqrt{n}$. The primes p with $p_m<p\le n$ are divided into $\varphi(A)$ residue classes $(\bmod\,A)$ with $\lambda_s n/\log n$ in each class $s=1,\dots,\varphi(A)$. Here $\lambda_1+\cdots+\lambda_{\varphi(A)}=1+o(1)$. Thus the differences p_i-p_j where p_i,p_j belong to the same residue class $(\bmod\,A)$ number

$$\left(\lambda_1^2 + \cdots + \lambda_{\varphi(A)}^2 + o(1)\right) \frac{n^2}{2\log^2 n} \geq \left(\frac{1}{\varphi(A)} + o(1)\right) \frac{n^2}{2\log^2 n} \cdot \tag{17}$$

Since the number of integers $\leq n$ which are divisible by A is $\leq n/A$ it follows from (17) that one of these integers has at least

$$\left(1 + o(1)\right) \frac{A}{\varphi(A)} \cdot \frac{n}{2\log^2 n} \tag{18}$$

representations (15). Now

$$\frac{A}{\varphi(A)} = \prod_{i < (1/2)\log n} \left(1 + \frac{1}{p_i - 1}\right) > c \log\log n\,. \tag{19}$$

By a more careful application of this method we can prove that for any monotonically increasing $f(n)$, with $f(n) \to \infty$ as slowly as we please, there is a sequence $l_n < f(n)\log n$ for which

$$F(n, l_n)\log^2 n/n \to \infty\,. \tag{20}$$

We cannot prove (20) if we only assume $l_n < c\log n$ for some fixed c.

P. Erdös and E. G. Straus, University of California, Los Angeles