ACTA ARITHMETICA
XXIX (1976)

A conjecture of Erdös in number theory

by

R. R. HALL (Heslington)

Introduction. Let k be a positive integer and $F(x, k)$ denote the number of positive integers $n < x$ which have a divisor in every residue class prime to k. Erdös [1] proved that for every fixed $\varepsilon > 0$, we have

$$F(x, k) = (1+o(1))x$$

provided

$$k < 2^{(1-\varepsilon)\log\log x}.$$

Erdös conjectured that the following stronger result holds: if c is any fixed real number and

$$k = 2^{\log\log x + (c+o(1))\sqrt{\log\log x}} \tag{1}$$

then

$$F(x, k) \sim \frac{x}{\sqrt{2\pi}} \int_c^\infty e^{-y^2/2}\, dy. \tag{2}$$

It is well known that if $\nu(n)$ denotes the number of distinct prime factors of n then

$$\operatorname{card}\left(n < x: \frac{\nu(n) - \log\log x}{\sqrt{\log\log x}} > c\right) \sim \frac{x}{\sqrt{2\pi}} \int_c^\infty e^{-y^2/2}\, dy; \tag{3}$$

moreover if $\psi(n) \to \infty$ arbitrarily slowly as $n \to \infty$ then for almost all n, we have

$$2^{\nu(n)} \leqslant \tau(n) \leqslant \psi(n) 2^{\nu(n)} \tag{4}$$

where $\tau(n)$ denotes the number of divisors of n. Certainly n is not counted by $F(x, k)$ if $\tau(n) < \varphi(k)$, and if we combine equations (1) to (4), we can say rather approximately that the assertion is that a number with sufficient divisors to go round will almost surely have one in every residue class prime to k.

In this paper I prove the following result in this direction:

THEOREM. *Let $\xi(k)\to 0$ arbitrarily slowly as $k\to\infty$. If k and x are related by* (1), *and if the interval*

$$(5)\qquad \left(1-\exp\left(-\xi(k)(\log k)^{3/4}\right),1\right)$$

is free of real zeros of the Dirichlet L-functions (mod k), *then* (2) *holds.*

The required zero-free interval (5) is wider than that established by Siegel's theorem, which would correspond to replacing the exponent $\frac{3}{4}$ of $\log k$ in (5) by 1. But a result of Page [4] gives the

COROLLARY. *The conjecture holds for almost all k. More precisely, for every fixed A the number of exceptional k's not exceeding K is $O(K/\log^A K)$.*

For Page's Lemma 8 states that there exists an absolute positive constant C_1 such that if $K\geqslant 2$, there is at most one primitive character χ^* with modulus $k^*\leqslant K$ such that $L(\sigma,\chi^*)$ vanishes for some σ satisfying

$$(6)\qquad \sigma > 1-C_1/\log K.$$

Thus if k is exceptional, either $\xi(k)(\log k)^{3/4}<\log(C_1^{-1}\log K)$ or there is a character (mod k) induced by χ^*, i.e. $k^*|k$. The former, small k's are negligible in number if $\xi(k)\to 0$ sufficiently slowly, and there are at most K/k^* multiples of k^* not exceeding K. By Siegel's theorem, $L(\sigma,\chi^*)=0$ implies that

$$\sigma<1-C_2(A)(k^*)^{-1/A},$$

where $C_2(A)>0$ and depends on A only. Combining this with (6) we obtain the corollary.

In the proof of the theorem to follow, O-constants and Vinogradov's $\ll$ are always uniform. The limiting process implied by the o-notation is as x and k tend to infinity; these are equivalent by (1).

Proof of the theorem. Since $k/\varphi(k)\ll\log\log k$, $\varphi(k)$ also satisfies (1). We need only consider the integers $n<x$ for which $\tau(n)\geqslant\varphi(k)$, and if we take $\psi(n)=\log\log n$ in (4), this implies that

$$\nu(n)\geqslant\log\log x+\left(c+o(1)\right)\sqrt{\log\log x}.$$

In view of (3), it will be necessary and sufficient to show that all but $o(x)$ of these numbers have a divisor in every residue class prime to k. Let $I(x)$ be the interval $(g,h]$, where

$$\log g=(\log\log x)^3,\qquad \log h=\frac{\log x}{\log g},$$

and let

$$f(n)=\prod_{p^a\| n}p^a,\qquad p\in I(x).$$

By the familiar variance method of Turán, $\nu(n)-\nu(f(n))$ has normal order $6\log\log\log n$. Hence we may assume that

$$(7)\qquad \log\log x+\left(c+o(1)\right)\sqrt{\log\log x}<\nu(f(n))<2\log\log x,$$

and we follow Erdös [1] in constructing the divisors of n in each residue class prime to k from prime factors of $f(n)$. We shall require:

LEMMA 1. *For each l prime to k we have*

$$\sum_{\substack{g<p\leqslant h\\ p\equiv l(\mathrm{mod}\,k)}}\frac{1}{p}=\frac{L}{\varphi(k)}\left(1-\chi_1(l)M+E(l)\right)$$

where

$$L=\int_g^h\frac{(1+\log y)}{y\log^2 y}\,dy,\qquad LM=\frac{1}{\beta}\int_g^h\frac{y^{\beta-1}(1+\log y)}{y\log^2 y}\,dy$$

and

$$|E(l)|\ll(\log\log x)^{-4}.$$

Here β is the unique Siegel zero (mod k) if such exists, that is $L(\beta,\chi_1)=0$, $\beta>1-C/\log k$. It is known that if C is a sufficiently small positive absolute constant there is at most one such zero, moreover χ_1 must be real and non-principal. We define $M=0$ when there is no such β.

The proof of this follows from the formula

$$\sum_{\substack{g<p\leqslant h\\ p\equiv l(\mathrm{mod}\,k)}}\frac{1}{p}=\frac{\psi(h,k,l)}{h\log h}-\frac{\psi(g,k,l)}{g\log g}-\int_g^h\psi(y,k,l)\,d\left(\frac{1}{y\log y}\right)+O\left(\frac{1}{\sqrt{g}}\right)$$

and Satz 7.3 (p. 136) of Prachar [5], which gives

$$\psi(y,k,l)=\frac{y}{\varphi(k)}-\frac{y^\beta}{\beta\varphi(k)}+O\left(ye^{-b\sqrt{\log y}}\right)$$

uniformly for $(l,k)=1$ and $k<\exp(a\sqrt{\log y})$, a and b being absolute positive constants. This condition holds for $y\geqslant g$ if x is sufficiently large.

We may assume that $f(n)$ is squarefree since the number of integers $n<x$ with a repeated prime factor in $I(x)$ is $O(x/g)=o(x)$. Hence we have

$$f(n)\leqslant h^{2\log\log x}\leqslant\sqrt{x}\quad\text{ if }\quad x\geqslant e^8,$$

in view of (7). Let $\sum'$ denote summation over squarefree $m\leqslant\sqrt{x}$, all of whose prime factors lie in $I(x)$, and which fail to have a divisor in every residue class prime to k, moreover which satisfy

$$(8)\qquad \log\log x+\left(c+o(1)\right)\sqrt{\log\log x}<\nu(m)<2\log\log x.$$

Then it will be sufficient to prove that

$$\sum_m{}' \sum_{\substack{n<x\\ f(n)=m}} 1 = o(x).$$

To estimate the inner sum, note that $n = mq$ where $q < x/m$ and has no prime factor in $I(x)$. Since $h < \sqrt{x} < x/m$, a theorem of van Lint and Richert [3] gives

$$\sum_{\substack{n<x\\ f(n)=m}} 1 \ll \frac{x}{m} \prod_{p\in I(x)} \left(1-\frac{1}{p}\right) \ll \frac{x\log g}{m\log h}$$

by Mertens' formula. Hence it will be enough to show that

(9) $$\sum_m{}' \frac{1}{m} = o\left(\frac{\log h}{\log g}\right).$$

Let $l_1, l_2, \ldots, l_t$ denote an arbitrary set of residue classes prime to k. We refer to these as a good set if the congruence:

$$l_1^{\varepsilon_1} l_2^{\varepsilon_2} \ldots l_t^{\varepsilon_t} \equiv h (\operatorname{mod} k), \quad \text{each } \varepsilon_j = 0 \text{ or } 1,$$

has a solution for every h prime to k, that is, as the ε_j's vary over their 2^t possible choices, the left hand side runs through every reduced residue class. If m has t (distinct) prime factors p_j such that $p_j \equiv l_j (\operatorname{mod} k)$ for $1 \leqslant j \leqslant t$, evidently m has a divisor in every residue class prime to k if $l_1, \ldots, l_t$ is a good set. Let $\sum_{(t)}$ denote summation over bad sets of t l_j's. Then

(10) $$\sum{}' \frac{1}{m} \leqslant \sum_t \frac{1}{t!} \sum_{(t)} \prod_{i=1}^{t} \Bigg(\sum_{\substack{g<p\leqslant h\\ p\equiv l_i (\operatorname{mod} k)}} \frac{1}{p} \Bigg)$$

where t runs through the possible values of $\nu(m)$, that is, the range given by (8). Notice that as we remarked at the beginning of the proof, $\varphi(k)$ satisfies (1), so that it will be sufficient to deal with the case

(11) $$\log\varphi(k) + o\big(\sqrt{\log\varphi(k)}\big) < t\log 2 < 3\log\varphi(k).$$

By Lemma 1, we have that

$$\prod_{i=1}^{t} \Bigg(\sum_{\substack{g<p\leqslant h\\ p\equiv l_i (\operatorname{mod} k)}} \frac{1}{p} \Bigg) = \frac{L^t}{\varphi^t(k)} (1-M^2)^{t/2} \left(\frac{1-M}{1+M}\right)^{u/2} \left(1 + O\left(\frac{1}{(\log\log x)^3}\right)\right)$$

where

$$u = \chi_1(l_1) + \chi_1(l_2) + \ldots + \chi_1(l_t),$$

so that

$$-t \leqslant u \leqslant t, \quad u \equiv t (\operatorname{mod} 2).$$

There are

$$\binom{t}{(t-u)/2} \left(\frac{\varphi(k)}{2}\right)^t$$

choices of the set $l_1, l_2, \ldots, l_t$ to give a fixed u, hence if we sum over all sets of t l_i's, we obtain

$$\sum \prod_{i=1}^{t} \Bigg(\sum_{\substack{g<p\leqslant h\\ p\equiv l_i (\operatorname{mod} k)}} \frac{1}{p} \Bigg)^2 \leqslant \frac{L^{2t}}{\varphi^t(k)} (1+M^2)^t \left(1 + O\left(\frac{1}{(\log\log x)^3}\right)\right).$$

By the Cauchy–Schwarz inequality, we may combine this with (10) and deduce that

(12) $$\sum{}' \frac{1}{m} \leqslant (1+M^2)^{t/2} \sum_t \frac{L^t}{t!} \left(\frac{1}{\varphi^t(k)} \sum_{(t)} 1\right)^{1/2}$$

$$\ll \delta^{1/2} (1+M^2)^{t/2} \frac{\log h}{\log g}$$

provided

(13) $$\sum_{(t)} 1 \leqslant \delta\varphi^t(k)$$

when t satisfies (11). Now we refer to Theorem 2 of Erdös and Rényi [2]. This implies that when

$$t\log 2 \geqslant \log\varphi(k) + 2\log\frac{1}{\delta} + \log\left(\frac{\log\varphi(k)}{\log 2}\right) + 5\log 2,$$

we have (13). By (11), we may infer that (13) holds provided we choose δ so that

(14) $$\log\frac{1}{\delta} = o\big(\sqrt{\log\varphi(k)}\big).$$

We require an estimate for M, defined in Lemma 1, and it is at this point that we use the hypothesis that the interval (5) is free of real zeros of L-functions $(\operatorname{mod} k)$. We have

$$LM \ll \int_{\log g}^{\infty} e^{-(1-\beta)v} \frac{dv}{v} \leqslant 1 + \log\left(\frac{1}{(1-\beta)\log g}\right).$$

Since $L \sim \log\log x \sim \log k/\log 2$, we have

$$M \ll \xi(k)(\log k)^{-1/4}, \quad M^2 t \ll \xi^2(k)(\log k)^{1/2}.$$

We select δ so that

$$\log\frac{1}{\delta} = 2M^2t+(\log k)^{1/3}$$

(the last term in case $M = 0$) and note that as $\xi(k)\to 0$ as x, and so $k\to\infty$, (14) is satisfied. With this choice of δ, (12) implies that (9) holds, and the proof is complete.

Remark. It may be that the Cauchy–Schwarz inequality is inefficient in deriving (12) and that the factor $(1+M^2)^{t/2}$ is not needed. However, I could not find a useful estimate for the number of bad sets $l_1, l_2, \ldots \ldots, l_t$ with a fixed u — evidently there is no uniformly good estimate of this type since if $u = t$ all the sets are bad, indeed

$$\chi_1(l_1^{\varepsilon_1}l_2^{\varepsilon_2}\ldots l_t^{\varepsilon_t}) = 1, \quad \varepsilon_j\text{'s arbitrary}.$$

Siegel's theorem gives the estimate $M = o(1)$ and it seems reasonable that rather more than this is needed.

References

[1] P. Erdös, *On the distribution of divisors of integers in residue classes* (mod d), Bull. Soc. Math. Grèce 6 (1) (1965), pp. 27–36.

[2] P. Erdös and A. Rényi, *Probabilistic methods in group theory*, J. Analyse Math. 14 (1965), pp. 127–138.

[3] J. H. van Lint and H.-E. Richert, *On primes in arithmetic progressions*, Acta Arith. 11 (1965), pp. 209–216.

[4] A. Page, *On the number of primes in an arithmetic progression*, Proc. London Math. Soc. 39 (1935), pp. 116–141.

[5] K. Prachar, *Primzahlverteilung*, Berlin 1957.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF YORK
Heslington, York, England

Received on 12. 7. 1974 (596)

ACTA ARITHMETICA
XXIX (1976)

On the equation of Catalan

by

R. TIJDEMAN (Leiden)

1. Introduction. The following conjecture was first enunciated by Catalan [8] in 1844 but has never been proved.

The only solution in integers $p > 1$, $q > 1$, $x > 1$, $y > 1$ of the equation

$$x^p - y^q = 1 \qquad (1)$$

is $p = y = 2$, $q = x = 3$.

In 1953 Cassels [6] independently made the weaker conjecture that equation (1) has only a finite number of solutions.

The equation has been shown to be impossible for some special values of p and q. In 1738 Euler [10] showed that the only solution of $x^2-y^3 = 1$ is $x = 3$, $y = 2$. In 1850 Lebesgue [14] proved that there is no solution at all when $q = 2$ and $p \neq 3$. It was shown by Nagell [18] in 1921 that there are no solutions if $p = 3$ or if $q = 3, p \neq 2$. The problem of showing that there is no solution when $p = 4$ was posed by Nagell and solved by S. Selberg [20] in 1932. Since 1967 this last result has become a special case of a theorem of Chao Ko [9], that there are no solutions if $p = 2$. Hence one has $p \geqslant 5$ and $q \geqslant 5$ for all unknown solutions of (1).

In proving Catalan's conjecture one can obviously assume without loss of generality that p and q are different primes. In 1960 Cassels [7] showed that if (1) holds then $p|y$ and $q|x$. It is an easy consequence of Cassels' result that there are no three consecutive positive integers which are all perfect powers, [17].

There are several results concerning the number of solutions when some of the variables are fixed. If x and y are fixed, then there are only finitely many solutions (p, q) of (1). This follows from Gel'fond's transcendence measure for $\log x/\log y$, [11]. LeVeque [15] showed that there is at most one solution (p, q) which can be found explicitly if it exists. Cassels [6] simplified his proof. If p and q are fixed, it is an immediate consequence of a result of Siegel [21] that (1) has only finitely many solutions (x, y). See also Mahler [16]. In this case Hyyrö [12] proved that there are at most $\exp(631p^2q^2)$ solutions. An explicit upper bound for