Can. J. Math., Vol. XXV, No. 6, 1973, pp. 1216–1223

ON THE DISTRIBUTION OF SQUARE-FREE NUMBERS

C. HOOLEY

1. Introduction. Erdös† [1] has shown that, if the square-free numbers in ascending order be denoted by $s_1, s_2, \ldots, s_n, \ldots$, then for $0 \leqq \gamma \leqq 2$

$$\sum_{s_{n+1} \leqq x} (s_{n+1} - s_n)^\gamma \sim B(\gamma)x$$

as $x \to \infty$. In this paper we shall extend this result by proving that the asymptotic formula in fact holds for the wider range $0 \leqq \gamma \leqq 3$.

Similar results have been obtained previously by the author in respect of both the sequence of numbers expressible as the sum of two squares and also sequences of numbers relatively prime to given large integers, although the method used here differs from that of the earlier papers [2; 3]. Our result may also be compared with the inequality

$$\sum_{p_{n+1} \leqq x} \frac{(p_{n+1} - p_n)^2}{p_n} = O(\log^3 x)$$

that has been obtained by Selberg on the Riemann hypothesis [6].

The method here admits of an immediate generalization for the treatment of k-free numbers, the corresponding asymptotic formula being stated without proof in Theorem 2 at the end of the paper. When $k \geqq 3$, however, it is possible to extend the analysis a little further in order to widen again the allowable range of γ in terms of k. Since it would not be appropriate to give the details here, an account of this development will be reserved for a future occasion.

2. Notation. The letters d, l, m, n, and v are positive integers; i, j, and μ are non-negative integers; p is a (positive) prime number.

The letter x denotes a variable that is to be regarded as tending to infinity, all appropriate inequalities that are true for sufficiently large x being assumed to hold.

The letter c with subscript is a positive absolute constant; the constants implied by the O-notation are absolute, being, in particular, independent of η in section 5; on the other hand, the passage to the limit as $x \to \infty$ implied by the o-notation is not necessarily uniform with respect to any other parameters involved.

Received June 22, 1972.

†Although Erdös only considers the case $\gamma = 2$, his method is immediately applicable to the case $0 \leqq \gamma \leqq 2$.

https://doi.org/10.4153/CJM-1973-129-0 Published online by Cambridge University Press

3. The initial inequality. Let $N_l = N_l(x)$ be the number of intervals $s_{n+1} - s_n$ of length l for which $s_{n+1} \leqq x$, and define the function $f_T(m)$ by

$$f_T(m) = \sum_{p^2 | m, p > c_1 T \log T} 1,$$

where $c_1 = 1/100$ and $T < x$ is any integer exceeding a sufficiently large positive absolute constant c_2. Then in this section we show that an inequality for a sum involving N_l can be expressed in terms of $f_T(m)$.

At the very beginning we follow Erdös [**1**] by considering, for any integer $l \geqq T$, the integers in any interval of the form $M < m < M + l$ that are divisible by the square of at least one prime p such that $p \leqq c_1 T \log T$. These integers do not in number exceed

$$\begin{aligned} \sum_{p \leqq c_1 T \log T} \left(\frac{l}{p^2} + 1\right) &= l \sum_{p \leqq c_1 T \log T} \frac{1}{p^2} + \pi(c_1 T \log T) \\ &< l\left(\frac{1}{4} + \sum_{r=2}^{\infty} \frac{1}{r(r+1)}\right) + \pi(c_1 T \log T) \\ &\leqq \frac{7}{8} l - 1. \end{aligned}$$

We deduce that, if $s_{n+1} - s_n = l$, then there are at least $l/8$ integers between s_n and s_{n+1} that have the property that they are divisible by the square of a prime p with $p > c_1 T \log T$, since all numbers between consecutive square-free numbers are divisible by the square of some prime.

Next we consider the $l - T$ intervals of the form $M \leqq m < M + T$ which consist entirely of integers lying between given consecutive square-free numbers differing by l. Then, for $l \geqq 32T$, at least $(l - T)/32$ of these intervals contain at least $T/32$ integers divisible by the square of a prime exceeding $c_1 T \log T$. For, were this not so, then at least $31(l - T)/32$ intervals would contain fewer than $T/32$ such integers, and it would follow that

$$\frac{31}{32}(l - T)\frac{1}{32} T + \frac{1}{32}(l - T)T \geqq \left(\frac{1}{8} l - 2T\right) T$$

as each of at least $l/8 - 2T$ of these integers is included in T different intervals of length T. But

$$\frac{1}{8} l - 2T \geqq \frac{1}{16}(l - T)$$

for $l \geqq 32T$, and the falsity of our assertion would therefore imply the impossible inequality $63/32^2 > 1/16$.

To relate this to the function $f_T(m)$ we define $F_T(n)$ by

$$F_T(n) = \sum_{n \leqq m < n+T} f_T(m),$$

and consider the sum

$$\sum_{n \leqq x-T} F_T^{\,2}(n). \tag{1}$$

https://doi.org/10.4153/CJM-1973-129-0 Published online by Cambridge University Press

Plainly, for $l \geqq 32T$ we have $F_T(n) \geqq T/32$ for at least $(l - T)/32 > l/64$ integers n such that both n and $n + T - 1$ lie between two consecutive square-free integers distance l apart. The sum (1) therefore counts at least $c_3 l T^2$ for each interval $s_{n+1} - s_n$ of length l, and we deduce that

$$c_3 T^2 M_{32T}(x) \leqq \sum_{n \leqq x-T} F_T^{\,2}(n), \tag{2}$$

where $M_v(x)$ is defined for any integer v by

$$M_v(x) = \sum_{l \geqq v} N_l\, l.$$

In the next section we pass on to the problem of obtaining an upper bound for the sum on the right-hand side of this inequality.

4. The sum $\sum_{n \leqq x-T} F_T^{\,2}(n)$. To estimate this sum we introduce the additional functions

$$g_X(m) = \sum_{p^2 | m, X < p \leqq 2X} 1$$

and

$$G_{T,X}(n) = \sum_{n \leqq m < n+T} g_X(m),$$

where throughout it will be assumed that $n \leqq x - T$ and

$$T < x^{\frac{1}{4} - \delta} \tag{3}$$

for some $\delta > 0$.

To find an inequality for $F_T(n)$ we write

$$F_T(n) = \sum_{X_i < x^{\frac{1}{2}},\, i \geqq 0} \frac{1}{X_i^{\frac{1}{4}} \log^{\frac{1}{4}} X_i} X_i^{\frac{1}{4}} \log^{\frac{1}{4}} X_i\, G_{T,X_i}(n),$$

where $X_i = c_1 2^i T \log T$, and then infer from the Cauchy-Schwarz inequality that

$$F_T^{\,2}(n) \leqq c_4 \left(\sum_{i=0}^{\infty} \frac{1}{2^{\frac{1}{2}i} T^{\frac{1}{2}} \log T} \right) \left(\sum_{X_i < x^{\frac{1}{2}},\, i \geqq 0} X_i^{\frac{1}{2}} \log^{\frac{1}{2}} X_i\, G_{T,X_i}^{\,2}(n) \right)$$

$$\leqq \frac{c_5}{T^{\frac{1}{2}} \log T} \sum_{X_i < x^{\frac{1}{2}},\, i \geqq 0} X_i^{\frac{1}{2}} \log^{\frac{1}{2}} X_i\, G_{T,X_i}^{\,2}(n).$$

Therefore

$$\sum_{n \leqq x-T} F_T^{\,2}(n) \leqq \frac{c_5}{T^{\frac{1}{2}} \log T} \sum_{X_i < x^{\frac{1}{2}},\, i \geqq 0} X_i^{\frac{1}{2}} \log^{\frac{1}{2}} X_i \sum_{n \leqq x-T} G_{T,X_i}^{\,2}(n). \tag{4}$$

We suppress temporarily the subscript from X_i in order to facilitate the consideration of the inner sum, although it will still be assumed that

https://doi.org/10.4153/CJM-1973-129-0 Published online by Cambridge University Press

$c_1 T \log T \leqq X < x^{\frac{1}{2}}$. We have

$$\sum_{n \leqq x-T} G_{T,X}^2(n) = \sum_{n \leqq x-T} \Big(\sum_{n \leqq m < n+T} g_X(m) \Big)^2$$
$$\leqq T \sum_{m \leqq x} g_X^2(m) + 2 \sum_{r<T} (T-r) \sum_{m \leqq x-r} g_X(m) g_X(m+r)$$
$$\leqq T \sum_{m \leqq x} g_X^2(m) + 2T \sum_{r<T} \sum_{m \leqq x-r} g_X(m) g_X(m+r)$$
$$= T(R(x, X) + 2S(x, X, T)), \text{ say.} \tag{5}$$

An estimate for $R(x, X)$ is easily derived. We have from the definition of $g_X(m)$ that

$$R(x, X) \leqq \sum_{m \leqq x} \Big(\sum_{\substack{p_1^2 | m, p_2^2 | m, \\ p_1, p_2 > X}} 1 \Big)$$
$$= \sum_{p > X} \left[\frac{x}{p^2} \right] + \sum_{\substack{p_1 \neq p_2, \\ p_1, p_2 > X}} \left[\frac{x}{p_1^2 p_2^2} \right]$$
$$< x \sum_{p>X} \frac{1}{p^2} + x \Big(\sum_{p>X} \frac{1}{p^2} \Big)^2 = O\Big(\frac{x}{X \log X} \Big). \tag{6}$$

An estimate for $S(x, X, T)$ could be obtained by following a similar method. Since this would not yield sufficiently precise bounds, we adopt instead a method that utilises an elementary principle in the theory of diophantine approximation. The sum $S(x, X, T)$ does not exceed the number of solutions in primes p_1, p_2 and integers l_1, l_2 of the inequality

$$0 < p_2^2 l_2 - p_1^2 l_1 < T$$

for which $p_2^2 l_2 < x$ and $X < p_1, p_2 \leqq 2X$. Since these conditions imply that $l_1, l_2 < x/X^2$ and $(l_1, l_2) < T$, we may write $l_1 = dl_1'$, $l_2 = dl_2'$, where $(l_1', l_2') = 1$ and $d < M_{X,T} = \min(x/X^2, T)$. Therefore, substituting for l_1, l_2 and then suppressing superscripts, we have

$$S(x, X, T) \leqq \sum_{d < M_{X,T}} V(x, X, T; d), \tag{7}$$

where $V(x, X, T; d)$ is the number of solutions in l_1, l_2, p_1, p_2 simultaneously satisfying the conditions

$$0 < p_2^2 l_2 - p_1^2 l_1 < T/d; \; p_2^2 l_2 < x/d, \tag{8}$$
$$(l_1, l_2) = 1, \tag{9}$$
$$X < p_1, p_2 \leqq 2X. \tag{10}$$

We shall find two estimates for $V(x, X, T; d)$, the estimate to be chosen for substitution in (7) depending on the value of d relative to x, X, and T.

Firstly we have

$$V(x, X, T; d) = \sum_{\substack{l_1, l_2 < x/dX^2 \\ (l_1, l_2) = 1}} V_{l_1, l_2}^{(1)}, \tag{11}$$

https://doi.org/10.4153/CJM-1973-129-0 Published online by Cambridge University Press

where $V_{l_1,l_2}^{(1)} = V_{l_1,l_2}^{(1)}(x, X, T; d)$ is, for given l_1, l_2, the number of solutions of (8) and (10) in primes p_1, p_2. For any such solution we have

$$0 < \frac{l_2}{l_1} - \frac{p_1^2}{p_2^2} < \frac{T}{dl_1p_2^2}$$

and therefore

$$0 < \left(\frac{l_2}{l_1}\right)^{\frac{1}{2}} - \frac{p_1}{p_2} < \frac{T}{dl_1p_1p_2} < \frac{T}{dl_1X^2}.$$

Hence, for given l_1, l_2, all ratios p_1/p_2 lie in a fixed interval of length not exceeding T/dl_1X^2. Also, because $X > T/d$ implies that $p_1 \neq p_2$, different solutions in p_1, p_2 always correspond to different ratios p_1/p_2. Since two different values of p_1/p_2 are at least $1/4X^2$ apart, we deduce that

$$V_{l_1,l_2}^{(1)} \leqq \frac{4T}{dl_1} + O(1),$$

and then from this and (11) that

$$V(x, X, T; d) = O\left(\frac{T}{d}\sum_{l_1,\, l_2 < x/dX^2} \frac{1}{l_1}\right) + O\left(\sum_{l_1,\, l_2 < x/dX^2} 1\right)$$

$$= O\left(\frac{Tx \log x}{d^2X^2}\right) + O\left(\frac{x^2}{d^2X^4}\right). \tag{12}$$

However, we also have that

$$V(x, X, T; d) = \sum_{X < p_1, p_2 \leqq 2X} V_{p_1,p_2}^{(2)}, \tag{13}$$

where $V_{p_1,p_2}^{(2)} = V_{p_1,p_2}^{(2)}(x, T; d)$ is, for given p_1, p_2, the number of solutions in l_1, l_2 of (8) and (9). Next, writing $\Lambda_j = 2^j$, we consider, for each j such that $\Lambda_j < x/dX^2$, only those solutions for which

$$\Lambda_j \leqq l_1 < 2\Lambda_j.$$

Any such solution satisfies

$$0 < \frac{l_2}{l_1} - \frac{p_1^2}{p_2^2} < \frac{T}{dl_1p^2} < \frac{T}{d\Lambda_jX^2},$$

and two different solutions give rise to two different ratios l_2/l_1 at least $1/4\Lambda_j^2$ apart. Hence the number of these solutions is at most

$$\frac{4T\Lambda_j}{dX^2} + O(1),$$

and therefore, summing over j, we have

$$V_{p_1,p_2}^{(2)} = O\left(\frac{Tx}{d^2X^4}\right) + O(\log x)$$

https://doi.org/10.4153/CJM-1973-129-0 Published online by Cambridge University Press

for $p_1, p_2 > X$. Finally from this and (13) we have

$$V(x, X, T; d) = O\left\{\left(\frac{Tx}{d^2X^4} + \log x\right) \sum_{X<p_1,p_2\leqq 2X} 1\right\}$$

$$= O\left(\frac{Tx}{d^2X^2\log^2 X}\right) + O\left(\frac{X^2\log x}{\log^2 X}\right), \tag{14}$$

which furnishes us with the second estimate for $V(x, X, T; d)$ that we require.

To decide which estimate for $V(x, X, T; d)$ should be used for each value of d in (7) we observe that it suffices to compare the second terms in the right-hand sides of (12) and (14), since the first terms in these are nearly the same. As these second terms are almost equal for $d = x/X^3$, the method of choosing between (12) and (14) and the consequent estimates for $S(x, X, T)$ are as follows.

(i) If $X \geqq x^{\frac{1}{3}}$, then we take (12) for all d in the summation. This gives

$$S(x, X, T) = O\left(\frac{Tx\log x}{X^2}\sum_{d=1}^{\infty}\frac{1}{d^2}\right) + O\left(\frac{x^2}{X^4}\sum_{d=1}^{\infty}\frac{1}{d^2}\right)$$

$$= O\left(\frac{Tx\log x}{X^2}\right) + O\left(\frac{x^2}{X^4}\right) \tag{15}$$

$$= O\left(\frac{x}{X}\right),$$

since $T < x^{\frac{1}{4}-\delta}$.

(ii) If $(x/T)^{\frac{1}{3}} \leqq X < x^{\frac{1}{3}}$, then we take (14) for $d \leqq x/X^3$ and (12) for $x/X^3 < d < T$, since $T < X < x/X^2$. Therefore

$$S(x, X, T) = O\left(\frac{Tx\log x}{X^2}\sum_{d=1}^{\infty}\frac{1}{d^2}\right) + O\left(X^2\sum_{d\leqq x/X^3} 1\right) + O\left(\frac{x^2}{X^4}\sum_{d>x/X^3}\frac{1}{d^2}\right)$$

$$= O\left(\frac{Tx\log x}{X^2}\right) + O\left(\frac{x}{X}\right) \tag{16}$$

$$= O\left(\frac{x}{X}\right),$$

since $X/T \geqq x^{4\delta/3}$.

(iii) If $c_1 T\log T \leqq X < (x/T)^{\frac{1}{3}}$, then we take (14) for $d < T$. In this case

$$S(x, X, T) = O\left(\frac{Tx}{X^2\log^2 X}\sum_{d=1}^{\infty}\frac{1}{d^2}\right) + O\left(\frac{X^2\log x}{\log^2 X}\sum_{d<T} 1\right)$$

$$= O\left(\frac{Tx}{X^2\log^2 X}\right) + O\left(\frac{TX^2\log x}{\log^2 X}\right) \tag{17}$$

$$= O\left(\frac{x}{X\log X}\right),$$

by a simple comparison of orders of magnitude.

Next by (6) we see that the estimates provided by (15), (16), and (17) hold in respect of not only $S(x, X, T)$ but also the coefficient of T in the right-hand side of (5). Therefore, by this and (4), we have

$$\begin{aligned}\sum_{n \leqq x-T} F_T^2(n) &= O\left(\frac{xT^{\frac{1}{2}}}{\log T}\left\{\sum_{X_i<(x/T)^{\frac{1}{3}}} \frac{1}{X_i^{\frac{1}{2}} \log^2 X_i} + \sum_{X_i \geqq (x/T)^{\frac{1}{3}}} \frac{\log^{\frac{1}{2}} X_i}{X_i^{\frac{1}{2}}}\right\}\right)\\ &= O\left(\frac{x}{\log^2 T} \sum_{i=0}^{\infty} \frac{1}{2^{\frac{1}{2}i}}\right) + O\left(\frac{T^{\frac{2}{3}} x^{\frac{5}{6}} \log x}{\log T} \sum_{i=0}^{\infty} \frac{i+1}{2^{\frac{1}{2}i}}\right) \qquad (18)\\ &= O\left(\frac{x}{\log^2 T}\right),\end{aligned}$$

which is the estimate required for the application of (2) in the coming section.

5. The theorem and its generalisation. To complete the proof of the theorem we require the following two lemmata, due, respectively, to Richert [**5**] and Mirsky [**4**].

LEMMA 1. *We have, for $s_{r+1} \leqq x$, that*

$$s_{r+1} - s_r = O(x^{2/9} \log x).$$

LEMMA 2. *For any given l we have*

$$N_l(x) = A(l)x + o(x).$$

Let γ be a real constant such that $0 \leqq \gamma \leqq 3$. Then, by Lemma 1, we have

$$\sum_l N_l l^\gamma = \sum_{l \leqq \eta} N_l l^\gamma + \sum_{\eta < l < u} N_l l^\gamma = \sum\nolimits_1 + \sum\nolimits_2, \text{ say,} \qquad (19)$$

where η is a given integer exceeding $32c_2$ and where $u = c_6 x^{2/9} \log x$ for some sufficiently large constant c_6. In this case

$$\sum\nolimits_1 = x \sum_{l \leqq \eta} A(l) l^\gamma + o(x) \qquad (20)$$

by Lemma 2. Next we infer from (2) and (18) that

$$M_v(x) = O\left(\frac{x}{v^2 \log^2 v}\right)$$

for $\eta < v < u$ on account of the inequalities satisfied by η and u. Therefore†

$$\begin{aligned}\sum\nolimits_2 &= O\left(\sum_{2^\mu \eta < u, \mu \geqq 0} M_{2^\mu \eta}(x) (2^\mu \eta)^{\gamma-1}\right)\\ &= O\left(x \sum_{\mu=0}^{\infty} \frac{1}{(2^\mu \eta)^{3-\gamma} \log^2 (2^\mu \eta)}\right) \qquad (21)\\ &= O\left(\frac{x}{\log \eta}\right).\end{aligned}$$

†Only the terms corresponding to $\mu = 0$ need be used for the case $0 \leqq \gamma \leqq 1$.

https://doi.org/10.4153/CJM-1973-129-0 Published online by Cambridge University Press

We conclude from (19), (20), and (21) that

$$\frac{1}{x}\sum_{l} N_l l^\gamma = \sum_{l \leqq \eta} A(l) l^\gamma + o(1) + O\left(\frac{1}{\log \eta}\right).$$

The final result is now almost immediate. We have

$$\left.\begin{array}{l} \overline{\lim}_{x\to\infty} \frac{1}{x}\sum_{l} N_l l^\gamma \\ \underline{\lim}_{x\to\infty} \frac{1}{x}\sum_{l} N_l l^\gamma \end{array}\right\} = \sum_{l \leqq \eta} A(l) l^\gamma + O\left(\frac{1}{\log \eta}\right)$$

from which firstly follows the convergence of $\sum_{l=1}^{\infty} A(l) l^\gamma$. Denoting the sum of the latter series by $B(\gamma)$, we then let $\eta \to \infty$ and obtain

$$\lim_{x\to\infty} \frac{1}{x}\sum_{l} N_l l^\gamma = B(\gamma).$$

We have thus proved the following theorem.

THEOREM 1. *Let the square-free numbers in ascending order be denoted by* $s_1, s_2, \ldots, s_n, \ldots$. *Then, for* $0 \leqq \gamma \leqq 3$, *we have*

$$\sum_{s_{n+1} \leqq x} (s_{n+1} - s_n)^\gamma \sim B(\gamma) x$$

as $x \to \infty$.

As stated in the introduction this theorem is a special case of the next theorem, which can be established by means of a routine generalisation of the above proof.

THEOREM 2. *Let the k-free numbers* $(k > 1)$ *in ascending order be denoted by* $s_1^{(k)}, s_2^{(k)}, \ldots, s_n^{(k)}, \ldots$. *Then, for* $0 \leqq \gamma \leqq k+1$, *we have*

$$\sum_{s_{n+1}^{(k)} \leqq x} (s_{n+1}^{(k)} - s_n^{(k)})^\gamma \sim B_k(\gamma) x$$

as $x \to \infty$.

REFERENCES

1. P. Erdös, *Some problems and results in elementary number theory*, Publ. Math. Debrecen *2* (1951), 103–109.
2. C. Hooley, *On the difference of consecutive numbers prime to n, I*, Acta Arith. *8* (1963), 343–347.
3. ——— On the intervals between numbers that are sums of two squares, *Acta Math. 127* (1971), 279–297.
4. L. Mirsky, *Arithmetical pattern problems relating to divisibility by r-th powers*, Proc. London Math. Soc. *50* (1949), 497–508.
5. H. E. Richert, *On the difference between consecutive square-free numbers*, J. London Math. Soc. *29* (1954), 16–20.
6. A. Selberg, *On the normal density of primes in small intervals and the difference between consecutive primes*, Archiv Math. og Naturvid. *47* (1943), 87–106.

University College, Cardiff,
Cardiff, Wales