PACIFIC JOURNAL OF MATHEMATICS
Vol. 59, No. 2, 1975

DISTRIBUTION OF SQUAREFREE INTEGERS IN NON-LINEAR SEQUENCES

IVAN E. STUX

I. This paper investigates the occurrences of the squarefree integers in sequences $s_n = [f(n)]$, $n = 1, 2, 3, \cdots$ where $f(x)$ belongs to classes of functions described by 'smoothness' conditions. The result obtained is an extention of the well known fact that $Q(x) = 6/\pi^2 x + O(x^{1/2})$, where $Q(x)$ = number of squarefree integers $\leqq x$; it states that $Q_s(x) \sim 6/\pi^2 g(x)$ where $Q_s(x)$ = number of squarefree integers $\leqq x$ in the sequence s_n, and $g(x)$ is the inverse function of $f(x)$.

This result relates to the deep theorem of Piateskii-Shapiro which states that if $1 < c < 12/11$ then the sequence $[n^c]$ has the proper rate of primes occuring, namely, $\pi_c(x) \sim x^{1/c}/\log x$.

The classes of functions used is described by the following:

DEFINITION 1. for given $1 < c < 2$, $0 < \delta < 1$

(1) $S(c, \delta)$ = set of functions $f(x)$ such that for some constant $a > 0$ depending on f, and for sufficiently large x's, depending on f,

$$(ax^c)^{(i)} \leqq (f(x))^{(i)} < (ax^{c+\delta})^{(i)}$$

holds for $i = 0, 1, 2$, the superscripts indicating the i^{th} derivative.

Functions like x^c, $1 < c < 2$, or more generally $\sum_{i=1}^{k} a_i x^{c_i} (\log x)^{d_i}$, where the leading term has $a > 0$, $1 < c < 2$, belong to these classes of functions.

The following theorem will be proved:

THEOREM 1. *Let $1 < c < 4/3$, then there exists a $\delta_c = \delta(c) > 0$, some small value depending on c such that if $f(x) \in S(c, \delta)$, $0 \leqq \delta < \delta(c)$, then*

$$Q_s(x) = 6/\pi^{-2} g(x) + O((g(x))^{1-\varepsilon}) \tag{2}$$

holds for some $\varepsilon > 0$ depending on c and δ, where $Q_s(x)$ = number of $\{s_n \leqq x, s_n = [f(n)], s_n = \text{squarefree}, n = 1, 2, 3, \cdots\}$, $g(x)$ = inverse function of f, $[z]$ = integer part of z.

II. Following are the lemmas that will be used in the proof:

LEMMA 1 (Piateskii-Shapiro, [2]). *Let $f(x)$ be a continuously*

$g(x)$ be its inverse function. Then, for an integer m such that $m = [f(n)]$, either $\{g(m)\} = 0$ or $1 - g'(m-1) < \{g(m)\} < 1$. Conversely, if $\{g(m)\} = 0$ or $1 - g'(m+1) < \{g(m)\} < 1$, then it follows that for some n, $m = [f(n)]$.

(The curly brackets indicate the fractional part of the real number, the straight brackets the integer part, as usual.)

LEMMA 2 (a theorem of Erdös-Turán, [1]). *If $\mu_1, \mu_2, \cdots$ is a real sequence and if D_N denotes its discrepancy modulo one, then for each integer $m \geqq 1$ we have*

$$ND_N \leqq K\left(\frac{N}{m+1} + \sum_{t=1}^{m} \frac{1}{t}\left|\sum_{n=1}^{N} e(t\mu_n)\right|\right) \tag{3}$$

(where K is a constant and $e(z) = e^{2\pi i z}$, as usual).

LEMMA 3 (Van der Corput, pg, 64, [3]). *Let $g(x)$ be a real function with a continuous and steadily decreasing derivative $g'(x)$ in (a, b), and let $g'(b) = \alpha$, $g'(a) = \beta$. Then*

$$\sum_{a<n\leqq b} e(g(n)) = \sum_{\alpha-\eta<\nu<\beta+\eta} \int_a^b e(g(x) - \nu x)dx + O(\log(\beta - \alpha + 2)) \tag{4}$$

where η is any positive constant less than one.

LEMMA 4 (Van der Corput, pg. 61, [3]). *Let $F(x)$ be a real function, twice differentiable, and let $F''(x) \geqq r > 0$, or $F''(x) \leqq -r < 0$ throughout the interval (a, b), then*

$$\left|\int_a^b e^{iF(x)}dx\right| \leqq \frac{8}{\sqrt{r}} . \tag{5}$$

III. The first part of the proof is aimed at establishing the uniform distribution modulo one and the discrepancy of that distribution for sequences $g(q)$ where q are squarefree integers and $g(x)$ is the inverse function of a function in $S(c, \delta)$ (where δ is usually small, depending on c). The following is the result in this direction:

THEOREM 2. *For given $1 < c < 2$, and $\delta > 0$, small enough depending on c alone, let $f(x) \in S(c, \delta)$ and let $g(x)$ be the inverse function of $f(x)$. Then the sequence $\{g(q): q \leqq K,\ q$ the squarefree integers$\}$ is uniformly distributed modulo one and*

$$N(K, \xi) = \xi Q(K) + Q(K)D_{Q(K)}(g) , \tag{6}$$

and

differentiable function with $f'(x) > 0$, $f''(x) \geqq 0$, *for* $x \geqq 1$, *and let*

$$Q(K)D_{Q(K)}(g) \ll K^{3/5+(c+2\delta)/(5c(c+\delta))} + K^{1-1/(2c(c+\delta))} \tag{7}$$

where $Q(K)$ = *number of squarefree integers* $\leqq K$, $N(K, \xi)$ = *number of elements in the sequence* $g(q)$, $q \leqq K$, q *squarefree, which fall into a fixed interval of length* $\xi\,(< 1)$ *modulo one, and* $D_{Q(K)}(g)$ *is the discrepancy, modulo one, of the sequence* $g(q)$.

Clearly, uniform distribution holds whenever $\delta > 0$ is small enough to make the exponents in the estimate (7) less than one.

Proof. For $h \geqq 1$, consider

$$T_h(K) = \sum_{\substack{q \leqq K \\ q \text{ squarefree}}} e(hg(q)) \,, \qquad e(z) = e^{2\pi i z} \,. \tag{8}$$

Suppose that K_0 is the large value from where on the estimates of g, g', g'' induced by the definition 1 hold, and let $K > K_0$, then

$$T_h(K) = \sum_{\substack{K_0 < q \leqq K \\ q \text{ squarefree}}} e(hg(q)) + O(K_0) \,, \tag{8'}$$

and

$$\sum_{\substack{K_0 < q \leqq K \\ q \text{ squarefree}}} e(hg(q)) = \sum_{\substack{K_0 < n \leqq K \\ n = \text{integer}}} e(hg(n)) \sum_{d^2 \mid (n, P^2)} \mu(d) \tag{9}$$

where $p = \Pi_{p \leqq K^{1/2}} p$, p = primes, (a, b) = greatest common divisor, $\mu(d)$ = Möbius function. We can further write

$$\begin{aligned} &\sum_{\substack{d \leqq \sqrt{K} \\ d \mid P}} \mu(d) \sum_{K_0/d^2 < m \leqq K/d^2} e(hg(d^2m)) \\ &\quad = \sum_{\substack{d \leqq A \\ d \mid P}} \mu(d) \sum_{K_0/d^2 < m \leqq K/d^2} e(hg(d^2m)) \\ &\qquad + \sum_{\substack{A < d \leqq K^{1/2} \\ d \mid P}} \mu(d) \sum_{K_0/d^2 < m \leqq K/d^2} e(hg(d^2m)) \,. \end{aligned} \tag{10}$$

We will pick the value of A later. The second sum in (10) can be estimated trivially as

$$\sum_{A < d \leqq K^{1/2}} \left| \sum_{K_0/d^2 < m \leqq K/d^2} e(hg(d^2m)) \right| \ll \sum_{A < d \leqq K^{1/2}} \frac{K}{d^2} \ll \frac{K}{A} \,. \tag{11}$$

The first sum, on the other hand, is estimated by

$$\sum_{d \leqq A} \left| \sum_{K_0/d^2 < m \leqq K/d^2} e(hg(d^2m)) \right| . \tag{12}$$

To estimate the inner sum, divide the interval $K_0/d^2 < m < K/d^2$ up into pieces of type $1/2^r\, K/d^2 < m \leqq 1/2^{r-1}\, K/d^2$, to get

$$(12')\qquad \sum_{d\leqq A}\sum_{r}\left|\sum_{\substack{1/2^r\,K/d^2<m\leqq 1/2^{r-1}\,K/d^2\\ m\geqq K_0/d^2}} e(hg(d^2m))\right|.$$

We will estimate the last inner sum by using Lemma 3 and then Lemma 4. The conditions in definition 1 give that

$$\left(\frac{y}{a}\right)^{(c+\delta)^{-1}} < g(y) \leqq \left(\frac{y}{a}\right)^{1/c},\quad \frac{1}{c+\delta}\left(\frac{y}{a}\right)^{(c+\delta)^{-1}-1} < g'(y) \leqq \frac{1}{c}\left(\frac{y}{a}\right)^{1/c},$$

the chain rule tells us that $(d/dx)g(d^2x) = [(d/dz)g(z)]\cdot d^2$, $z = d^2x$, and so we have, by Lemma 3, for each r

$$(13)\qquad \sum_{1/2^r\,K/d^2<m\leqq 1/2^{r-1}\,K/d^2} e(hg(d^2m)) = \sum_{\nu} I_\nu + E\,,$$

where the $\sum_\nu$ extends over $(1/a_1(c+\delta))h((1/2^{r-1})K)^{(c+\delta)^{-1}-1}\,d^2 - 1/2 < \nu < (1/a_2c)h((1/2^r)K)^{c^{-1}-1}d^2 + 1/2$, $a_1 = a^{(c+\delta)^{-1}}$, $a_2 = a^{1/c}$, and

$$(14)\qquad I_\nu = \int_{2^{-r}Kd^{-2}}^{2^{-r+1}Kd^{-2}} e(hg(d^2x) - \nu x)dx\,,$$

and

$$E = O(\log(\max\nu - \min\nu + 2))\,.$$

In (14), first we change variables to $y = d^2x$, and then apply Lemma 4

$$(14')\qquad I_\nu = \frac{1}{d^2}\int_{2^{-r}K}^{2^{-r+1}K} e\left(hg(y) - \frac{\nu y}{d^2}\right)dy$$

but here $d^2/dy^2(hg(y) - \nu y/d^2) \geqq (c+\delta)^{-1}((c+\delta)^{-1} - 1)(1/a)hy^{(c+\delta)^{-1}-2}$. thus, we get, applying Lemma 4 that

$$(15)\qquad I_\nu \ll \frac{1}{d^2}\left[h\left(\frac{1}{2^{r-1}}K\right)^{(c+\delta)^{-1}-2}\right]^{-1/2}.$$

We thus have for (12) the estimate:

$$(16)\qquad \ll \sum_{d\leqq A}\sum_{r}\sum_{\nu}\frac{1}{d^2}h^{-1/2}\frac{2^{(r-1)/(c+\delta)c}}{2^{r-1}}K^{1-1/(2c+2\delta)} + \sum_{d\leqq A}\sum_{r}(E)$$

(for largest r we might get a shorter range of integration in (14), but the upper bound estimates still clearly hold in (16)). where $\sum_\nu$ is over

$$\frac{1}{a(c+\delta)}h\left(\frac{1}{2^{r-1}}K\right)^{(c+\delta)^{-1}-1}d^2 - \frac{1}{2} < \nu < \frac{1}{ac}h\left(\frac{1}{2^r}K\right)^{c^{-1}-1}d^2 + \frac{1}{2}\,.$$

From here, we have that the ν summation is bounded by

$$\ll hd^2 2^r\left(\frac{1}{2^{r-1}}\right)^{1/(c+\delta)}K^{c^{-1}-1} + 1$$

and so we can further estimate (12) by

$$(16') \quad \begin{aligned} &\ll \sum_{d \leqq A} \sum_{r} \left(hd^2 \frac{2^r}{2^{(r-1)/(c+\delta)}} K^{(1/c)-1} + 1 \right)\left(\frac{h^{-1/2}}{d^2} \frac{2^{(r-1)/(2(c+\delta))}}{2^{r-1}} K^{1-1/(2c+2\delta)} \right) \\ &+ \sum_{d \leqq A} \sum_{r} E \ll Ah^{1/2}K^{c^{-1}-1/(2c+2\delta)} + h^{-1/2}K^{1-1/(2c+2\delta)} + A \log K \,. \end{aligned}$$

(The last step is because $\sum_r$ in the first term was just a geometric sum and so it converges, while in the second part, the number of terms of the $\sum_r$ is $O(\log K)$.) The estimate (16′) together with (11) now gives us that

$$(17) \qquad T_h(K) \ll Ah^{1/2}K^{(c+2\delta)/(c+\delta)2c} + h^{-1/2}K^{1-1/(2c+2\delta)} + \frac{K}{A} + K_0 \,.$$

Here, for K sufficiently large the last error term absorbs into the first one if $A \geqq 1$ (which will anyway be the case). We now pick A so as to balance the 1st and 3rd terms of (17), i.e. let $A = [K^{2^{-1}-(c+2\delta)/(c+\delta)4c}h^{-1/4}]$. With this choice we obtain

$$(18) \qquad T_h(K) \ll h^{1/4}K^{2^{-1}+(c+2\delta)/(c+\delta)4c} + h^{-1/2}K^{1-1/(2c+2\delta)} \,.$$

Finally we use Lemma 2 to write:

$$(19) \quad \begin{aligned} Q(K)D_{Q(K)}(g) &\ll \frac{K}{m+1} + \sum_{h=1}^{m} \frac{1}{h}\{h^{1/4}K^{2^{-1}+(c+2\delta)/(c+\delta)4c} + h^{-1/2}K^{1-1/(2c+2\delta)}\} \\ &\ll \frac{K}{m+1} + m^{1/4}K^{2^{-1}+(c+2\delta)/(c+\delta)4c} + K^{1-1/(2c+2\delta)} \,. \end{aligned}$$

We pick the optimal m, i.e. $m = [K^{4(2^{-1}-(c+2\delta)/(c+\delta)4c)/5}]$, and thus we have

$$(20) \qquad Q(K)D_{Q(K)}(g) \ll K^{3/5+(c+2\delta)/(c+\delta)4c} + K^{1-1/(2c+2\delta)} \,.$$

COROLLARY 1. *If $1 < c < 4/3$ then there exists $\delta_c > 0$ depending on c such that if $f(x) \in S(c, \delta)$ for $0 < \delta < \delta_c$, and $g(x)$ is its inverse function then:*

$$(21) \qquad Q(K)D_{Q(K)}(g) \ll K^{(c+\delta)^{-1}-\varepsilon} \ll (g(K))^{1-\varepsilon}$$

for some $\varepsilon > 0$, depending on c and δ.

Proof. All we need to show is that

$$\frac{3}{5} + \frac{c+2\delta}{5c(c+\delta)} < \frac{1}{c+\delta} \quad \text{and} \quad 1 - \frac{1}{2(c+\delta)} < \frac{1}{c+\delta}$$

hold for some $\delta > 0$. By continuity it is enough to check that $3/5 + 1/5c < 1/c$ and $1 - 1/2c < 1/c$ hold. But the first of these holds if

$c < 4/3$, the second if $c < 3/2$.

IV. We can now prove Theorem 1. Let

$$T_s(x, y) = \text{number of } \{s_n = [f(n)],\ y < s_n \leqq x,\ s_n = \text{squarefree},\ n = 1, 2, 3, \cdots\} \tag{22}$$

Clearly, $T_s(1, y) = Q_s(y)$. Lemma 1 can now be used together with expressions (6) and (21). ξ in (6) will be taken $g'(y+1)$ or $g'(x-1)$ to give upper and lower bounds on $T(x, y)$, where $g(x)$ is as usual the inverse function of $f(x)$. We obtain:

$$T_s(x, y)\begin{cases} < g'(x-1)(Q(y) - Q(x)) + O(y^{(c+\delta)^{-1}-\varepsilon}) \\ > g'(y+1)(Q(y) - Q(x)) + O(y^{(c+\delta)^{-1}-\varepsilon}) \end{cases} \tag{23}$$

where $Q(x) = \#$ squarefree integers $\leqq x$. Or

$$\begin{aligned} &T_s(x, y) \\ &\begin{cases} < g'(x)(Q(y) - Q(x)) + O(y^{(c+\delta)^{-1}-\varepsilon}) + O(x^{c^{-1}-2}(Q(y) - Q(x))) \\ > g'(y)(Q(y) - Q(x)) + O(y^{(c+\delta)^{-1}-\varepsilon}) + O(y^{c^{-1}-2}(Q(y) - Q(x))) \,. \end{cases} \end{aligned} \tag{24}$$

Thus, for $0 < \alpha < 1$, using the well-known fact that $Q(x) = 6\pi^{-2}x + O(x^{1/2})$,

$$T_s(x, (1+\alpha)x) = \begin{cases} < \dfrac{6}{\pi^2}\, x \cdot \alpha \cdot g'(x) \\ > \dfrac{6}{\pi^2}\, x \cdot \alpha \cdot g'((1+\alpha)x) \end{cases} + O(x^{(c+\delta)^{-1}-1-\varepsilon} + x^{1/2}) \,. \tag{25}$$

On the other hand, clearly

$$Q_s(x) = \sum_{k=1}^{L(x)} T_s\Big(\frac{x}{(1+\alpha)^k},\ \frac{x}{(1+\alpha)^{k-1}}\Big) + O(1) \tag{26}$$

holds for an appropriate function $L(x)$ which tends to ∞ for $x \to \infty$, if $\alpha = \alpha(x) > 0$ is some given function of x which tends to zero as $x \to \infty$ (the relation is $(1 + \alpha(x))^{L(x)} \cong x$).

Using (25) in the expression (26) we obtain

$$Q_s(x)\begin{cases} < \dfrac{6}{\pi^2}\displaystyle\sum_{k=1}^{L(x)} \frac{x \cdot \alpha}{(1+\alpha)^k}\, g'\Big(\frac{x}{(1+\alpha)^k}\Big) + O(L(x) \cdot x^\gamma) \\ > \dfrac{6}{\pi^2}\displaystyle\sum_{k=1}^{L(x)} \frac{x \cdot \alpha}{(1+\alpha)^k}\, g'\Big(\frac{x}{(1+\alpha)^{k-1}}\Big) + O(L(x) \cdot x^\gamma) \end{cases} \tag{27}$$

where $\gamma = \max\{1/(c+\delta) - \varepsilon, 1/2\}$ and so it is actually $1/(c+\delta) - \varepsilon$. The main terms of the expressions on the right of (27) are exactly the upper and lower approximating sums of the Riemann integral

$$\int_1^x g'(y)dy = g(x) - g(1) .$$

To see how closely these sums approximate the integral, it suffices to find out how closely they are to each other, i.e. to estimate:

$$\begin{aligned}
\Delta(x) &= \Big|\sum_{k=1}^{L(x)} \frac{x\alpha}{(1+\alpha)^k} g'\Big(\frac{x}{(1+\alpha)^{k-1}}\Big) \\
&\quad - \sum_{k=1}^{L(x)} \frac{x\cdot\alpha}{(1+\alpha)^k} g'\Big(\frac{x}{(1+\alpha)^{k}}\Big)\Big| \\
&= \Big|\sum_{k=0}^{L(x)-1} \frac{x\cdot\alpha}{(1+\alpha)^{k+1}} g'\Big(\frac{x}{(1+\alpha)^{k}}\Big) \\
&\quad - \sum_{k=1}^{L(x)} \frac{x\cdot\alpha}{(1+\alpha)^k} g'\Big(\frac{x}{(1+\alpha)^{k}}\Big)\Big| \qquad (28) \\
&\leqq \Big|\alpha\cdot\sum_{k=1}^{L(x)-1} \frac{x}{(1+\alpha)^k}\Big(\frac{1}{1+\alpha} - 1\Big) g'\Big(\frac{x}{(1+\alpha)^{k}}\Big)\Big| \\
&\quad + |x\cdot\alpha\cdot g'(x)| + O(\alpha) \\
&\ll \Big|\alpha\cdot\sum_{k=1}^{L(x)-1} \frac{x\cdot\alpha}{(1+\alpha)^{k+1}} g'\Big(\frac{x}{(1+\alpha)^{k}}\Big)\Big| \\
&\quad + O\,|x\cdot\alpha\cdot g'(x)| + O(\alpha) .
\end{aligned}$$

The last sum is now the lower estimating sum of the integral, so one can write for $\alpha = \alpha(x)$

$$\Delta(x) \ll \alpha(x)(g(x) - g(1)) + \alpha(x)\cdot x\cdot g'(x) , \qquad (29)$$

so

$$\Delta(x) \ll \alpha(x)g(x) + O(1) . \qquad (30)$$

Equation $(1+\alpha(x))^{L(x)} \cong x$ gives us that $\alpha(x)$ and $L(x) = [\log x/\alpha(x)]$ is a pair for which expression (26) holds; picking in particular $\alpha(x) = (\log x(g(x))^{-\varepsilon})^{1/2}$, gives

$$\Delta(x) \ll \sqrt{\log x}\,(g(x))^{1-\varepsilon/2} , \qquad (31)$$

and

$$L(x)\cdot x^{\gamma} \ll L(x)(g(x))^{1-\varepsilon} = \sqrt{\log x}\,(g(x))^{1-\varepsilon/2} .$$

Calling ε' some value $0 < \varepsilon' < \varepsilon/2$ yields

$$Q_s(x) = \frac{6}{\pi^2} g(x) + O((g(x)^{1-\varepsilon'}) . \qquad (32)$$

I take this opportunity to express my appreciation to Professor Harold N. Shapiro for his continued advice and valuable suggestions.

References

1. P. Erdös and P. Turán, *On a problem in the theory of uniform distribution*, Indagationes Math., **10** (1948), 370–378, 406–412.
2. I. I. Piateskii-Shapiro, *On the distribution of prime numbers in sequences of form* $[n^c]$, Math. Sbornik, **33** (1953).
3. E. C. Titchmarsh, *The Theory of the Riemann Zeta Function*, Oxford University Press, (1951).

Received June 30, 1975. This work was partially supported by the National Science Foundation, Grant NSF-GP-33019X.

COLUMBIA UNIVERSITY, NEW YORY CITY