AVERAGES OF EXPONENTS IN FACTORING INTEGERS¹

IVAN NIVEN

For any positive integer \( m > 1 \), say \( m = p_1^{a_1}p_2^{a_2} \cdots p_r^{a_r} \), define \( h(m) = \min (a_1, a_2, \cdots, a_r) \) and \( H(m) = \max (a_1, a_2, \cdots, a_r) \). For convenience take \( h(1) = 1 \) and \( H(1) = 1 \). We prove that

\[
\begin{align*}
(1) \quad \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} h(j) &= 1, \\
(2) \quad \lim_{n \to \infty} \frac{1}{n} \sum_{j=1}^{n} H(j) &= 1 + \sum_{k=2}^{\infty} \left\{ 1 - \frac{1}{\zeta(k)} \right\},
\end{align*}
\]

where \( \zeta(k) \) is the zeta function. The infinite series on the right side of (2) converges to 0.7 approximately.

We prove something more than (1). Paul Erdös suggested to me in correspondence that it was likely that

\[
\sum_{j=1}^{n} h(j) = n + c\sqrt{n} + o(\sqrt{n}).
\]

His conjecture turned out to be correct, with \( c = \zeta(3/2)/\zeta(3) \). Formula (3) with this value of \( c \) is proved in §1, and of course this also establishes (1). We prove (2) in §2. In §3 we discuss normal order of \( h(n) \) and \( H(n) \).

1. Proof of (3). Let \( S \) be the set of squares of the natural numbers, and let \( S(n) \) denote the number of elements of \( S \) that do not exceed \( n \). Let \( T \) be the set of positive integers \( m \) such that \( h(m) \geq 2 \). Thus \( T \supseteq S \), and any element \( m \) of \( T \) not in \( S \) can be written uniquely in the form

\[
m = k^2q_1q_2 \cdots q_t, \quad (q_1q_2 \cdots q_t) | k, \quad t \geq 1,
\]

where \( q_1, q_2, \cdots, q_t \) are distinct primes. Now fix \( q_1q_2 \cdots q_t \) and consider the number of elements of \( T \) that are \( \leq n \) and have the form (4) for some \( k \). This is the same as the number of squares \( \leq n/(q_1q_2 \cdots q_t) \) that are divisible by \( q_1q_2 \cdots q_t \). Now for any real \( x > 0 \) the number of positive squares \( \leq x \) that are divisible by \( q_1q_2 \cdots q_t \) is \( \lesssim (x/(q_1q_2 \cdots q_t)^t) \). Hence the number of elements of \( T \) that are \( \leq n \) and have the form (4) is

Received by the editors June 18, 1968.

¹ Research supported in part by NSF Grant GP-6510.
AVERAGES OF EXPONENTS IN FACTORIZING INTEGERS

\[ S(n/(q_1q_2\cdots q_t)^3). \]

Also we observe that
\[ \sqrt{x} - 1 < \lfloor \sqrt{x} \rfloor = S(x) \leq \sqrt{x}. \]

Thus if we sum the terms (5) over all subsets \( q_1, q_2, \ldots, q_t \) of \( p_1, p_2, \ldots, p_r \), where \( p_r \) is the \( r \)th prime and \( p_{r+1} > n \), we see that
\[ T(n) = \sum S(n/(p_1^\beta_1 p_2^\beta_2 \cdots p_r^\beta_r)) \]
where the sum is over the \( 2^r \) terms with each \( \beta_i = 0 \) or 3. This gives
\[ T(n) \leq \sum S\left( \frac{n}{p_1^{\beta_1} p_2^{\beta_2} \cdots p_r^{\beta_r}} \right) \leq \sqrt{n} \prod_{i=1}^r (1 + p_i^{-3/2}) \]
by use of (6). But also for \( s > 1 \) we have \( \xi(s) = \prod (1 - p_i^{-s})^{-1} \) where the product is over all primes, and so
\[ T(n)/\sqrt{n} \leq \xi(3/2)/\xi(3). \]

On the other hand if we choose \( N > (p_1 p_2 \cdots p_r)^3 \) we can rewrite (7) as an inequality with \( n \) replaced by \( N \),
\[ T(N) > S(N/(p_1^\beta_1 p_2^\beta_2 \cdots p_r^\beta_r)) > \sqrt{N} \prod_{i=1}^r (1 + p_i^{-3/2}) - 2^r \]
by use of (6). It follows that
\[ T(N)/\sqrt{N} > \xi(3/2)/\xi(2) \prod_{i> r} (1 + p_i^{-3/2})^{-1} - 2^r/N. \]

By taking \( r \) large enough, then \( N \) large enough, we see that this with (8) implies
\[ \lim_{n \to \infty} T(n)/\sqrt{n} = \xi(3/2)/\xi(3), \]
\[ T(n) = \sqrt{n} \xi(3/2)/\xi(3) + o(\sqrt{n}). \]

Next let \( S_3 \) denote the set of cubes of the natural numbers, and let \( T_3 \) denote the positive integers \( m \) such that \( h(m) \geq 3 \). Then \( S_3 \subset T_3 \), and any member \( m \) of \( T_3 \) not in \( S_3 \) can be written uniquely in the form
\[ m = k q_1 a_1 q_2 a_2 \cdots q_t a_t, \quad (q_1 q_2 \cdots q_t) | k, \quad t \geq 1, \quad a_i = 1 \text{ or } 2, \]
where \( q_1, q_2, \ldots, q_t \) are distinct primes. Now fix \( q_1^{a_1} q_2^{a_2} \cdots q_t^{a_t} \) and consider the number of integers \( \leq n \) that belong to \( T_3 \) and have the form (10) for some \( k \). The number of such integers is, by an argument analogous to that leading to (5),
Then the analog to (7) is

\[(12) \quad T_3(n) = S_3(n/(p_1^{\beta_1} p_2^{\beta_2} \cdots p_r^{\beta_r}))\]

where the sum is over the \(3^r\) terms with each \(\beta_i = 0, 4, \text{ or } 5, \) and \(r\) is chosen so that \(p_{r+1} > n. \) Now \(S_3(x) \leq x\) so that (12) leads to

\[
\begin{align*}
T_3(n) & \leq \sum \left\{ \frac{n}{p_1^{\beta_1} p_2^{\beta_2} \cdots p_r^{\beta_r}} \right\}^{1/3} \\
& < n^{1/3} \prod_p (1 + p^{-4/3}) \prod_p (1 + p^{-5/3}) \\
& = n^{1/3} \zeta(4/3) \zeta(5/3) \zeta(8/3) \zeta(10/3)^{-1}, \\
T_3(n) & = O(n^{1/3}).
\end{align*}
\]

A similar argument leads to \(T_k(n) = O(n^{1/k})\) for any integer \(k \geq 3,\) where \(T_k\) denotes the positive integers \(m\) such that \(H(m) \geq k.\) However, we need this only for \(k = 3\) in this paper.

Now we examine the positive integers \(h(1), h(2), \cdots, h(n).\) The number of these that exceed 1 is \(T(n);\) the number that exceed 2 is \(T_3(n).\) Also we note that

\[
\max\{h(1), h(2), \cdots, h(n)\} = \lfloor \log_2 n \rfloor.
\]

It follows that

\[
n + T(n) \leq h(1) + h(2) + \cdots + h(n) \leq n + T(n) + T_3(n) \cdot \log_2 n.
\]

This with (9) and (13) establishes (3) with \(c = \zeta(3/2)/\zeta(3).\)

2. Proof of (2). Let \(Q_k\) be the set of \(k\)-power free integers, i.e. the set of positive integers \(m\) such that \(H(m) \leq k - 1.\) Now the number of integers \(m\) satisfying \(1 \leq m \leq n\) and \(H(m) = k - 1\) is \(Q_k(n) - Q_{k-1}(n).\) Also for \(n \geq 2\) the maximum of \(H(1), H(2), \cdots, H(n)\) is \(\lfloor \log_2 n \rfloor.\) Thus we see that

\[
\sum_{i=1}^{n} H(i) = \sum_{k=2}^{j+1} (k - 1)\{Q_k(n) - Q_{k-1}(n)\}, \quad j = \lfloor \log_2 n \rfloor.
\]

But \(Q_{j+1}(n) = n,\) so this can be written as

\[
\sum_{i=1}^{n} H(i) = jn - \sum_{k=2}^{j} Q_k(n), \quad j = \lfloor \log_2 n \rfloor.
\]
Next we prove that if \( r \) satisfies \( p_r^k \leq n \leq p_{r+1}^k \) then

\[
Q_k(n) = \sum (-1)^{a_1+a_2+\cdots+a_r} \left\lfloor \frac{n}{(p_1^{a_1} p_2^{a_2} \cdots p_r^{a_r})} \right\rfloor
\]

where the sum contains \( 2^r \) terms obtained by taking each \( \alpha_i = 0 \) or \( k \).

This can be established by interpreting \( \left\lfloor n/s \right\rfloor \) as the number of integers \( \leq n \) that are divisible by \( s \). Thus the right side of (15) can be interpreted as the number of integers from 1 to \( n \), first with those divisible by \( p_1^k \) deleted, then with those divisible by \( p_2^k \) deleted, \( \cdot \cdot \cdot \), then with those divisible by \( p_r^k \) deleted, then with those divisible by \( p_1^k p_2^k \) counted back in, and so on. Hence (15) follows by use of the inclusion-exclusion principle.

Next, equation (15) can be written as

\[
Q_k(n) = \sum \mu(d) \left\lfloor \frac{n}{d^k} \right\rfloor,
\]

where the sum is over all positive divisors \( d \) of \( p_1 p_2 \cdots p_r \). In this sum any term for which \( d^k > n \) has \( \left\lfloor n/d^k \right\rfloor = 0 \) so we can take the sum in (16) over all positive integers \( d \) satisfying \( d^k \leq n \). It is well known that

\[
\zeta(k)^{-1} = \prod_p \left( 1 - p^{-k} \right) = \sum_{d=1}^{\infty} \frac{\mu(d)}{d^k}.
\]

This with (16) gives

\[
n_{k}^{\zeta(k)^{-1}} - Q_k(n) = \sum_{d=1}^{\infty} \frac{\mu(d) \cdot n}{d^k} - \sum_{d^k \leq n} \mu(d) \left\lfloor \frac{n}{d^k} \right\rfloor = \sum_{d^k \leq n} \mu(d) \left\{ \frac{n}{d^k - \left\lfloor \frac{n}{d^k} \right\rfloor} \right\} + \sum_{d^k > n} \mu(d) \cdot \frac{n}{d^k},
\]

(17) \[ n_{k}^{\zeta(k)^{-1}} - Q_k(n) \] is less than \( n^{1/k} \). As to the second sum, its first term is less than 1, and all other terms are bounded by \( n \int_{n^{1/k}}^{\infty} x^{-k} dx = n^{1/k} \cdot ( -1 ) < n^{1/k} \)

since \( k \geq 2 \). Hence (17) can be written

\[
\left| n_{k}^{\zeta(k)^{-1}} - Q_k(n) \right| \leq n^{1/k} + 1 + n^{1/k} < 3n^{1/k}.
\]

This with (14) implies
\[ \left| n^{-1} \sum_{i=1}^{n} H(i) - 1 - \sum_{k=2}^{j} \{1 - \frac{1}{\zeta(k)-1}\} \right| \\
= \left| j - n^{-1} \sum_{k=2}^{j} Q_k(n) - 1 - \sum_{k=2}^{j} \{1 - \frac{1}{\zeta(k)-1}\} \right| \\
= n^{-1} \left| \sum_{k=2}^{j} \{n^{\frac{1}{\zeta(k)}} - Q_k(n)\} \right| \\
\leq n^{-1} \sum_{k=2}^{j} \left| n^{\frac{1}{\zeta(k)}} - Q_k(n) \right| \\
\leq n^{-1} \sum_{k=2}^{j} 3n^{1/k} \leq n^{-1}(3n^{1/2}) \log_2 n. \]

As \( n \) tends to infinity so does \( j = [\log_2 n] \), and hence we have (2).

3. **The normal order of \( h(n) \) and \( H(n) \).** We use the definition of normal order given in [1, p. 356]. It is clear from equation (1) that \( h(n) \) has normal order 1, since \( h(n) = 1 \) for almost all positive integers \( n \). On the other hand \( H(n) \) has no normal order, because it is an integer-valued function whose average order, by equation (2), lies between 1 and 2.

**Reference**


**University of Oregon**