ON POSITIVE INTEGERS $\leq x$ WITH PRIME FACTORS $\leq t \log x$

Andrew Granville
Department of Mathematics
University of Toronto
Toronto, Canada, M5S 1A1

ABSTRACT. It is not difficult to estimate the function $\psi(x,y)$, which counts integers $\leq x$, free of prime factors $> y$, by "smooth" functions whenever $y \leq \log^{1/2} x$ or y is a fixed power of x. This can be extended to $y < \log^{3/4} x$, and $y > \log^{2+\epsilon} x$ under the assumption of the Riemann Hypothesis. The real difficulty lies when y is a fixed multiple of $\log x$ and, in this paper, we investigate the set of integers $\leq x$, free of prime factors $> t \log x$, by estimating various functions related to $\psi(x, t\log x)$.

1. INTRODUCTION.

Define $S(x,y)$ to be the set of positive integers $\leq x$, composed only of prime factors $\leq y$. The cardinality of this set, $\psi(x,y)$, is called the Dickman-De Bruijn function and has been extensively investigated by many authors (see [14] for a review). In this section we will give some well-known results about $\psi(x,y)$ and sketch proofs of smooth asymptotic estimates when $y < \log^{1/2} x$ and when y is a fixed power of x. We also indicate how, in the literature, these have been extended to $y < \log^{3/4} x$, and to $y > \log^{2+\epsilon} x$ under the assumption of the Riemann Hypothesis.

It has not yet been possible to get a "smooth" asymptotic estimate for $\psi(x,y)$ when y is a fixed multiple of $\log x$ though, as we shall see, Hildebrand and Tenenbaum [11] have recently given an estimate in terms of certain functions of prime numbers; however at y around $\log x$, these functions seem to be very dependant on the "local" distribution of primes and so probably can't be estimated by a "smooth" function. Moreover, as has been noticed by many authors (and as we shall indicate in this section), the actual behaviour of $\psi(x,y)$ changes quite drastically at y around $\log x$.

In this paper we will investigate the set $S(x, t\log x)$ for fixed $t > 0$ as $x \to \infty$ by estimating the order of magnitude of various functions related to $\psi(x,y)$. In

Section 2 we introduce these functions and give short proofs of the estimates when $y/\log x \to 0$ or ∞ as $x \to \infty$.

When $y = t \log x$ such estimates are more difficult to prove, though by using a method of Erdös [5] together with some elementary combinatorial arguments we do this in Sections 3 and 4. Finally, in Section 5, we show that the question of estimating the order of magnitude of the subset of k-free integers in $S(x, t\log x)$, for arbitrary t and k, is equivalent to a notoriously difficult problem in entropy theory.

We start with an argument, due to Ennola [4], for estimating $\psi(x,y)$ when y is small:

Let $p_1 < p_2 < \cdots < p_n$ denote the primes up to y. It is clear that $p_1^{a_1} p_2^{a_2} \cdots p_n^{a_n} \leq x$ if and only if $a_1 \log p_1 + a_2 \log p_2 + \cdots + a_n \log p_n \leq \log x$. Therefore $\psi(x,y)$ is precisely the number of (integer) lattice points inside the n-dimensional tetrahedron defined by the bounding hyperplanes:

$$X_1 = 0, \quad X_2 = 0, \ldots, X_n = 0. \tag{1}$$

and

$$X_1 \log p_1 + X_2 \log p_2 + \cdots + X_n \log p_n \;=\; \log x. \tag{2a}$$

Thus $\psi(x,y)$ is the volume of the shape S given by constructing a unit box to the right and above each such lattice point. It is easy to see that the tetrahedron defined by (1) and (2a) lies completely inside S, which itself lies completely inside the tetrahedron defined by (1) and

$$(X_1 - 1)\log p_1 + \cdots + (X_n - 1)\log p_n = \log x. \tag{2b}$$

Thus we have the bounds

$$1 \leq \psi(x,y) / \frac{1}{\pi(y)!} \prod_{p \leq y} \frac{\log x}{\log p} \leq (1+\gamma)^{\pi(y)}, \tag{3}$$

where $\gamma = \theta(y)/\log x$ and $\theta(y) = \sum_{p\leq y} \log p$, as usual.

Now, by the prime number theorem, $(1+\gamma)^{\pi(y)} = 1 + O(\frac{y^2}{\log x \log y})$ whenever $y < \log^{1/2} x$; and so, by (3), we have the estimate

$$\psi(x,y) = \frac{1}{\pi(y)!} \prod_{p \leq y} \frac{\log x}{\log p} \{1 + O(\frac{y^2}{\log x \log y})\} \tag{4}$$

throughout the range

$$2 \leq y \leq \log^{1/2} x. \tag{5}$$

Unfortunately the bounds in (3) certainly do not give us a good estimate for $y > \log^{1/2+\epsilon} x$, though Ennola [4] managed to prove a similar type of formula to (4) for $y < \log^{3/4} x$, but by a different argument.

When y is large compared to x one can approach the question of estimating $\psi(x,y)$ by a totally different method. We sketch a proof of Dickman's result [3] that for any fixed $u > 0$,

$$\psi(x,y) \sim x\rho(u) \qquad (\text{as } x \to \infty,\ y = x^{1/u}), \tag{6}$$

where $\rho(u)$ is the continuous solution of the differential difference equation

$$u\rho'(u) + \rho(u-1) = 0 \qquad (u > 1), \tag{7}$$

with initial values

$$\rho(u) = 1 \qquad (0 < u \leq 1). \tag{8}$$

We first prove that (6) holds for $0 < u \leq 2$: if $u \leq 1$ then $\psi(x,y) = [x]$ for $y = x^{1/u}$. For $1 \leq u \leq 2$, we have $\rho(u) = 1 - \log u$ by (7) and

$$\begin{aligned}\psi(x,y) &= [x] - \sum_{x \geq p > y} [\frac{x}{p}] \\ &= x(1 - \sum_{x \geq p > y} \frac{1}{p}) + O(\pi(x)) \\ &= x(1 - \log u) + O(\frac{x}{\log x})\end{aligned}$$

by the prime number theorem.

Now let U be the set of values of u for which (6) fails and choose $u \in U$ within $1/2$ of the infimum of U, if U is non-empty. By the "Buchstab identity",

$$\psi(x, x^{1/v}) = \psi(x, x^{1/u}) + \sum_{x^{1/v} \geq p > x^{1/u}} \psi(\frac{x}{p}, p) \tag{9}$$

for $v = u - 1$, we see that

$$\psi(x, x^{1/u}) = x(\rho(v) - \sum_{x^{1/v} \geq p > x^{1/u}} \frac{1}{p}\rho(\frac{\log x}{\log p} - 1))\{1 + o(1)\} \tag{10}$$

as $\frac{\log x/p}{\log p} < \frac{\log x}{\log x^{1/u}} - 1 = v <$infimum U.

Now, by using a Riemann-Stieltjes integral we see that

$$\begin{aligned}\sum_{x^{1/v} \geq p > x^{1/u}} \frac{1}{p}\rho\left(\frac{\log x}{\log p} - 1\right) &= \int_{t=u}^{v} \frac{1}{x^{1/t}}\rho(t-1)d\pi(x^{1/t}) \\ &= \int_{t=v}^{u} \frac{\rho(t-1)}{t}dt + o(1)\end{aligned}$$

by the prime number theorem,

$$= \rho(v) - \rho(u) + o(1) \qquad \text{by (7)},$$

and (6) follows from substituting this into (10).

By replacing (9) with the identity

$$\psi(x,y)\log x = \int_1^x \frac{\psi(t,y)}{t}dt + \sum_{\substack{p^m \le x \\ p \le y}} \psi(\frac{x}{p^m}, y)\log p,$$

Hildebrand [9] has shown that one may extend the estimate (6) in the form

$$\psi(x,y) = x\rho(u)(1 + O_\epsilon(\frac{u\log(u+1)}{\log x})) \tag{6'}$$

for the range $x \ge y = x^{1/u} \ge 2$, $y \ge \log^{2+\epsilon} x$ where $\varepsilon > 0$ is fixed, under the assumption of the Riemann Hypothesis.

We've seen that it's possible to obtain good estimates for $\psi(x,y)$ by "smooth" functions whenever $y \ge \log^{2+\epsilon} x$ or $y < \log^{3/4} x$. However, it has proved difficult to do this in the range

$$\log^{3/4} x < y < \log^{2+\epsilon} x, \tag{11}$$

and it may be impossible (Hildebrand [9] conjectured that (6) does not hold for $y < \log^{2+\epsilon} x$). We review what is known:

In 1938 Rankin [15] gave the following simple but effective method to find an upper bound: Fix any $\sigma > 0$. Then

$$\psi(x,y) \le \sum_{\substack{n \ge 1 \\ P(n) \le y}} (\frac{x}{n})^\sigma = x^\sigma \zeta(\sigma, y) \tag{12}$$

where

$$\zeta(s,y) = \prod_{p \le y}(1 - p^{-s})^{-1}.$$

By elementary calculus we can see that in order to minimize the right hand side of (12) we should pick $\sigma = \alpha(x,y)$ to be the solution of

$$\log x = \sum_{p \le y} \frac{\log p}{p^{\alpha(x,y)} - 1}.$$

Actually Hildebrand and Tenenbaum [11] recently gave the estimate

$$\alpha(x,y) = \frac{\log(1 + y/\log x)}{\log y}\{1 + O(\frac{\log\log(1+y)}{\log y})\}, \tag{13}$$

throughout the range

$$x \geq y \geq 2. \tag{14}$$

(An estimate given, but not proved, by De Bruijn [1])

In fact Hildebrand and Tenenbaum succeeded in estimating $\psi(x,y)$ throughout the range (14) though, in terms of functions of prime numbers (such as in the right hand side of (12)) which are difficult to estimate in terms of smooth functions. Essentially they showed that one can estimate $\psi(x,y)$ asymptotically, by dividing $x^\alpha \zeta(\alpha, y)$ through by

$$\begin{array}{ll} \sqrt{2\pi u}\ \log(\frac{y}{\log x}) & \text{when}\quad \frac{y}{\log x} \to \infty \\ \sqrt{2\pi u}(1+\frac{1}{t})\ \log(1+t) & \text{when}\quad y = t\log x \\ \sqrt{2\pi y/\log y} & \text{when}\quad \frac{y}{\log x} \to 0 \end{array} \tag{15}$$

as $x \to \infty$ where $u = \log x/\log y$. Note how the behaviour of $\psi(x,y)/x^\alpha\zeta(\alpha,y)$ is quite different depending on the behaviour of $y/\log x$ as $x \to \infty$.

Actually this difference in behaviour, depending on the ratio $y/\log x$, is easily seen in the following theorem of De Bruijn [1], which measures only the order of magnitude of $\psi(x,y)$:

Theorem A The estimate

$$\log\ \psi(x,y) \sim \frac{\log x}{\log y}\ \log(1+\frac{y}{\log x}) + \frac{y}{\log y}\ \log(1+\frac{\log x}{y}) \tag{16}$$

holds uniformly in the range

$$x \geq y \geq 2, \qquad y = x^{1/u}. \tag{17}$$

We define, for each $t > 0$,

$$F(t) = \log(1+t) + t\ \log(1+1/t).$$

Then Theorem A may be stated as

Theorem 1 The estimate

$$\log\ \psi(x,y) \sim uF(y/\log x) \tag{18}$$

holds uniformly in the range (17).

This estimate was derived in an elementary way by Erdös and Van Lint [6] who gave the more aesthetically pleasing formula

$$\psi(x,y) \sim \binom{u+\pi(y)}{u}^{1+o(1)} \tag{19}$$

throughout the range (17).
Remark 1: It is important to note that the ratio $y/\log x \sim \pi(y)/u$ as $y \to \infty$.
Remark 2: In [14] (pp. 20-21) Norton criticizes both of the results (16) and (19) given above, suggesting that neither hold for fixed u. Of course this is quite untrue for (16), and also for (19) if one allows $u! = \Gamma(u+1)$ for any positive real u.

Hildebrand and Tenenbaum [11] account for the change in behaviour of $\psi(x,y)$ at y around $\log x$, as follows: If $y > y(\varepsilon)$ and $ky \le (1-\varepsilon)\log x$ then $(\prod_{p\le y} p)^k \le x$ by the prime number theorem, and so the numbers that do not have a prime factorization with exceptionally high powers contribute little to $\psi(x,y)$; whilst this feature does not occur when $y/\log x$ is large.

From this it seems logical to expect that the behaviour of the distribution of $\Omega(n)$ (the number of prime factors of n, counting repetitions) in the set $S(x,y)$ would change as y varies around $\log x$. However, Hildebrand [10] has shown that if $3 \le y \le \exp(\log^{1/21} x)$ then the distribution of $\Omega(n)$ on the set $S(x,y)$ is roughly Gaussian with mean u and standard deviation $V = \frac{u^{1/2}}{min(\log u, \log y)}(1 + \frac{\log x}{y})^{1/2}$.

$$(\text{i.e. } \frac{1}{\psi(x,y)} \sum_{\substack{m\in S(x,y)\\ \Omega(m)\le t}} 1 \;\approx\; \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{(t-u)/V} e^{-s^2/2} ds).$$

This distribution does not display a significant change in character as the ratio $y/\log x$ grows from 0 to ∞, (i.e. the distribution of $\Omega(n)$ depends more on u than $\pi(y)$) and so we must search for other explanations!

2. THE MAIN RESULTS

In this section we examine a number of other functions that count certain aspects of the set $S(x,y)$. In particular these functions are very "sensitive" as $y/\log x$ grows from 0 to ∞. Here we shall estimate these functions when $y/\log x \to 0$ as $x \to \infty$ and when $y/\log x \to \infty$ as $x \to \infty$, and in the next section we investigate them when y is a fixed multiple of $\log x$.

The first function that we consider is $\psi^{(2)}(x,y)$ which counts the number of squarefree integers in $S(x,y)$: Fix $\varepsilon > 0$.

If
$$y_0 \le y \le (1-\varepsilon)\log x \tag{20}$$
then $\prod_{p\le y} p \le x$ by the prime number theorem, and so, throughout the range (20),
$$\psi^{(2)}(x,y) = 2^{\pi(y)}. \tag{21}$$
By Theorem 1 this gives
$$\psi^{(2)}(x,y) = \psi(x,y)^{o(1)} \tag{22}$$

uniformly when

$$x \geq y \geq 2, \qquad y/\log x \to 0 \text{ as } x \to \infty. \tag{23}$$

For large y it is well known that $\psi^{(2)}(x,x) \sim x/\zeta(2)$ and Ivič and Tenenbaum [12] have shown that the estimate

$$\psi^{(2)}(x,y) = \frac{\psi(x,y)}{\zeta(2\alpha)}\{1+o(1)\}. \tag{24}$$

holds uniformly when

$$x \geq y \geq \log^{2+\epsilon} x \tag{25}$$

where $\alpha = \alpha(x,y)$ is given by (13).

We have the trivial bound $\psi^{(2)}(x,y) \leq \psi(x,y)$ and, when $y/\log x \to \infty$ and $u \to \infty$ as $x \to \infty$,

$$\begin{aligned}\psi^{(2)}(x,y) \geq\ & \#\{\text{ choices of } [u] \text{ distinct primes } \leq y\} = \binom{\pi(y)}{[u]} \\ \geq\ & \binom{\pi(y)+u}{u}^{1+o(1)} = \psi(x,y)^{1+o(1)} \qquad \text{by (19).}\end{aligned}$$

Combining this with (24) for when u is bounded we have the estimate

$$\psi^{(2)}(x,y) = \psi(x,y)^{1+o(1)} \tag{26}$$

uniformly in the range

$$x \geq y \geq 2, \qquad y/\log x \to \infty \quad \text{as} \quad x \to \infty. \tag{27}$$

Comparing (22) and (26) in the ranges (23) and (27) respectively we see that the value of $\log(\psi^{(2)}(x,t\log x))/\log(\psi(x,t\log x))$ goes from 0 to 1 (as $x \to \infty$) as t ranges from 0 to ∞. This is precisely the sort of "threshold function" that we have been looking for!

Define

$$F^{(2)}(t) = \begin{cases} t\log 2 & t \leq 2 \\ F(t\text{-}1) & t \geq 2 \end{cases}$$

We shall show

Theorem 2 The estimate

$$\log \psi^{(2)}(x,y) \sim uF^{(2)}(y/\log x) \tag{28}$$

holds uniformly in the range (17).

In Figure 1, we give the graph of

$$\frac{F^{(2)}(t)}{F(t)} = \lim_{x\to\infty} \frac{\log(\psi^{(2)}(x,t\log x))}{\log(\psi(x,t\log x))} \text{ as } t \text{ ranges from 0 to } \infty.$$

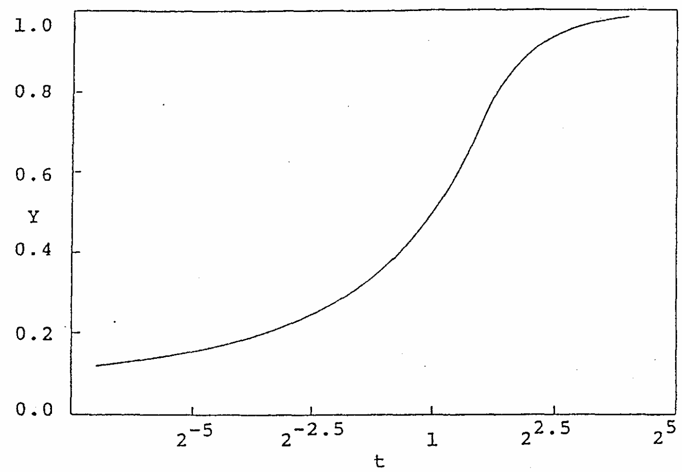

Figure 1: $Y = F^{(2)}(t)/F(t)$

In Section 5 we generalize this approach and try to estimate the order of magnitude of the function $\psi^{(k)}(x,y)$ (for each $k \geq 2$), which counts the number of k-free integers in $S(x,y)$. It turns out to be very difficult when y is around $\log x$.

The second function that we examine is $\psi(x,x',y)$, which counts the number of ordered pairs (m,n) of coprime integers where $m \in S(x,y)$ and $n \in S(x',y)$. This was introduced by Gunderson [8] in his Ph.D. Thesis, who gave an explicit lower bound similar to (4). Elsewhere [7] we will show that in the range (5), with $x \geq x'$ we have the uniform estimate, for $n = \pi(y)$,

$$\psi(x,x',y) = \frac{1}{n! \prod_{p \leq y} \log p} \sum_{j=0}^{n} \binom{n}{j}^2 \log^j x \log^{n-j} x' \{1 + O(\frac{y^2}{\log x \log y})\}$$

In particular, if $x = x'$ then by (4) we have

$$\psi(x,x,y) \sim \binom{2\pi(y)}{\pi(y)} \psi(x,y) \tag{29}$$

in the range (5).

Now if the pair (a,b) is counted by $\psi(x,x,y)$ then $n = ab \in S(x^2,y)$. But any such integer n can be factored in at most $2^{\pi(y)}$ distinct ways into an ordered pair of coprime integers (a,b), and so $\psi(x,x,y) \leq 2^{\pi(y)}\psi(x^2,y)$. But then, as $\psi(x,x,y) \geq \psi(x,y)$ we see from Theorem 1 that

$$\psi(x,x,y) = \psi(x,y)^{1+o(1)} \tag{30}$$

in the range (23).

For large y it is well known that $\psi(x,x;x) \sim x^2/\zeta(2)$ as $x \to \infty$, and, by using methods similar to Ivič and Tenenbaum [12], one can show that the estimate

$$\psi(x,x',y) = \frac{\psi(x,y)\psi(x'y)}{\zeta(\alpha+\alpha')}\{1 + O_\epsilon(\frac{1}{\log\, x'} + \frac{\log\log(y+1)}{\log\, y})\}$$

holds uniformly in the range

$$x \geq x' \geq 2, \qquad y \geq (\log x \log x')^{1+\epsilon},$$

where $\alpha = \alpha(x,y)$, $\alpha' = \alpha(x',y)$.

In particular, taking $x = x'$ we have

$$\psi(x,x,y) = \frac{\psi(x,y)^2}{\zeta(2\alpha)}\{1 + o(1)\} \tag{31}$$

uniformly in the range (25).

Of course $\psi(x,x,y) \leq \psi(x,y)^2$ and

$$\begin{aligned}\psi(x,x,y) &\geq \#\{\text{pairs of disjoint subsets A and B of } \{p_1,\ldots,p_n\} \text{ of size } \lfloor u \rfloor\} \\ &\geq \binom{\pi(y)}{2\lfloor u \rfloor}\binom{2\lfloor u \rfloor}{\lfloor u \rfloor} \\ &\geq \psi(x,y)^{2+o(1)} \quad \text{as} \quad \pi(y)/u \to \infty \text{ and } u \to \infty,\end{aligned}$$

by (19). Combining this with (31) for when u is bounded we have the estimate

$$\psi(x,x,y) = \psi(x,y)^{2+o(1)} \tag{32}$$

in the range (27) and so the value of $\log(\psi(x,x,t\log x))/\log(\psi(x,t\log x))$ (as $x \to \infty$) goes from 1 to 2 as t ranges from 0 to ∞.

Define

$$F_2(t) = \log((1+t^2)^{1/2} + t) + t\ \log((1+1/t^2)^{1/2} + 1/t).$$

We shall show

Theorem 3 The estimate

$$\log\ \psi(x^{1/2}, x^{1/2}, y) \sim uF_2(y/\log\, x)$$

holds uniformly in the range (17).

In Figure 2 we give the graph of

$$\frac{2F_2(t/2)}{F(t)} = \lim_{x\to\infty} \frac{\log\ \psi(x,x,t\ \log\ x)}{\log\ \psi(x,t\ \log\ x)} \text{ as } t \text{ ranges from 0 to } \infty.$$

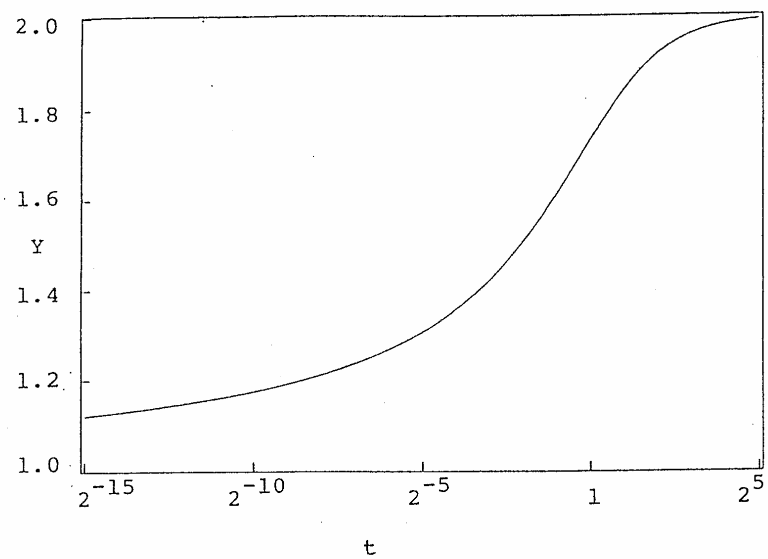

Figure 2: $Y = 2F_2(t/2) \ / \ F(t)$

If we look at the function $\psi(x^2,y)$ then we see, from Theorem 1, that $\psi(x^2,y) \sim \psi(x,y)^{1+o(1)}$ in the range (23), and $\psi(x^2,y) \sim \psi(x,y)^{2+o(1)}$ in the range (27), as $x \to \infty$ and so, from (30) and (32), it makes sense to compare $\psi(x,x,y)$ with $\psi(x^2,y)$ rather than with $\psi(x,y)$. By Theorems 1 and 3 we see that

$$\psi(x^2,y) \sim \psi(x,x,y)^{1+o(1)} \quad \text{if } y/\log x \to 0 \text{ or } \infty$$

and that if

$$G(t) = \lim_{x\to\infty} \frac{\log(\psi(x^{1/2},x^{1/2},t\log x))}{\log(\psi(x,t\log x))}$$

then $G(t) = F_2(t)/F(t)$. Thus $G(1/t) = G(t)$ and so we get a function that reaches its maximum at $t = 1$, which is symmetric about $t = 1$ (logarithmically). In Figure

3 we present the graph of $G(t)$.

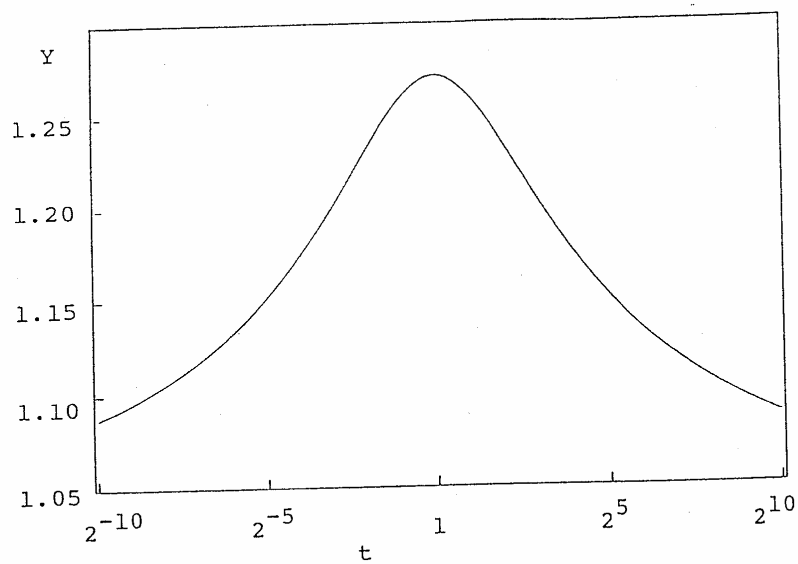

Figure 3: Y = G(t)

Evidently this graph is very interesting as the main "difference" between $\psi(x^{1/2}, x^{1/2}, y)$ and $\psi(x, y)$ seems to occur at $y \approx \log x$. To interpret this we may write

$$\psi(x,y) = \sum_{n \in S(x,y)} 1$$

and

$$\psi(x^{1/2}, x^{1/2}, y) = \sum_{n \in S(x,y)} \#\{ab = n : \ (a,b) = 1 \text{ and } a, b \leq x^{1/2}\}.$$

The graph in Figure 3 tells us that for an 'average' element n of $S(x,y)$, the number of pairs of coprime divisors of n, both less than $x^{1/2}$, is significant when compared to the cardinality of $S(x,y)$, only when y is a multiple of $\log x$; most significant when $y = \log x$.

Given this interpretation it seems logical to investigate

$$\begin{aligned}\psi^*(x,y) &= \sum_{n\in S(x,y)} \#\{ab=n:(a,b)=1\}\\ &= \sum_{n\in S(x,y)} 2^{\omega(n)}.\end{aligned}$$

Now

$$\psi(x,y)\le\psi^*(x,y)\le 2^{\omega}\psi(x,y)$$

where

$$\omega=\max_{n\in S(x,y)}\omega(n)\le\min\{\pi(y),(1+o(1))\frac{\log x}{\log\log x}\}$$

by the prime number theorem; and so $2^{\omega}=\psi(x,y)^{o(1)}$ by Theorem 1. Therefore $\psi^*(x,y)=\psi(x,y)^{1+o(1)}$ in the ranges (23) and (27), and we shall show

Theorem 4 The estimate

$$\log\ \psi^*(x,y)\sim uF_2(y/\log x)$$

holds uniformly in the range (17).

Together with Theorem 3 this gives

$$\log\ \psi(x^{1/2},x^{1/2},y)\sim\log\ \psi^*(x,y)$$

throughout (17); a result which is not too surprising given the respective definitions of these functions. The final function that we consider is the next step up from $\psi^*(x,y)$, namely

$$\begin{aligned}\psi^+(x,y) &= \sum_{n\in S(x,y)} \#\{ab=n\}\\ &= \sum_{n\in S(x,y)} \tau(n)\end{aligned}$$

where $\tau(n)$ is the divisor function. In this case we have

$$\psi(x,y)\le\psi^+(x,y)\le\tau\psi(x,y)$$

where

$$\tau=\max_{n\in S(x,y)}\tau(n)\le\min\{\psi(x,y),2^{(1+o(1))\frac{\log x}{\log\log x}}\}.$$

The second of the bounds is due to Wigert [16] and implies that $\psi^+(x,y) = \psi(x,y)^{1+o(1)}$ in the range (27); the first comes from observing that any divisor of an element of $S(x,y)$ must itself be an element of $S(x,y)$. We thus have

$$\psi^+(x,y) \le \psi(x,y)^2 \tag{33}$$

in the range (23); and so, as

$$\begin{aligned}\psi(x^{1/2},y)^2 &= \sum_{n\in S(x,y)} \#\{ab = n : a,b \le x^{1/2}\}\\ &\le \psi^+(x,y),\end{aligned}$$

we have, from Theorem 1, that

$$\psi^+(x,y) = \psi(x,y)^{2+o(1)}$$

in the range (23). We shall prove

Theorem 5 The estimate

$$\log\ \psi^+(x,y) \sim uF(2y/\log x)$$

holds uniformly in the range (17).

The pairs of functions $\psi^+(x,y)$ and $\psi(x^{1/2},y)^2$, and $\psi^*(x,y)$ and $\psi(x^{1/2},x^{1/2},y)$ have been seen to be closely related. We make the following:

Conjecture We have

$$1 \le \frac{\psi^+(x,y)}{\psi(x^{1/2},y)^2} \le \log^{1+o(1)} x$$

and

$$1 \le \frac{\psi^*(x,y)}{\psi(x^{1/2},x^{1/2},y)} \le \log^{1+o(1)} x$$

uniformly throughout (17).

Note that the lower bounds are both trivial, and that an upper bound of $0(\log x)$ is obtained when $y = x$. It seems plausible that the first of these two may succumb to the saddle point method employed in [11], though the second seems to be much more difficult.

3. THE METHOD OF PROOF WHEN y IS A FIXED MULTIPLE OF $\log x$.

We generalize a method of proof, used by Erdös in [5], when he first estimated $\log \psi(x,y)$ for y a fixed multiple of $\log x$. Let $R(x,y)$ represent $\psi(x,y)$, $\psi^{(2)}(x,y)$, $\psi(x^{1/2},x^{1/2},y)$, $\psi^*(x,y)$ or $\psi^+(x,y)$, according to which theorem we're proving! Fix $y = t\log x$, $z = y/\log y$ and $v = \log x/\log\log x$. We define $R(x,(z,y])$ as the function we had before, but only summing over integers composed of prime factors from the interval $(z,y]$. We have the inequality

$$R(x,y) \le R(x,z)\ R(x,(z,y]). \tag{34}$$

Now define

$$\begin{aligned}
M(v,n) &= \sum_{a_1+\cdots+a_n\le v} 1,\\
M^{(2)}(v,n) &= \sum_{\substack{a_1+\cdots+a_n\le v\\ \text{each } a_i\le 1}} 1,\\
M(\frac{v}{2},\frac{v}{2},n) &= \sum_{\substack{a_1+\cdots+a_n\le v/2\\ b_1+\cdots+b_n\le v/2}\ a_ib_i=0 \text{ for all } i} 1,\\
M^*(v,n) &= \sum_{a_1+\cdots+a_n\le v} 2^{\#\{i:a_i\ne 0\}}
\end{aligned}$$

and

$$M^+(v,n) = \sum_{a_1+\cdots+a_n\le v} (a_1+1)(a_2+1)\dots(a_n+1),$$

where in each sum the a_i's are non-negative integers.

If $p_1 < \cdots < p_n$ are primes then it is clear that

(i) If $a_1+\cdots+a_n \le \frac{\log x}{\log p_n}$ then $p_1^{a_1}p_2^{a_2}\cdots p_n^{a_n} \le x$;

and

(ii) If $p_1^{a_1}\cdots p_n^{a_n} \le x$ then $a_1+\cdots+a_n \le \frac{\log x}{\log p_1}$.

If we let $S(v,n)$ represent $M(v,n)$, $M^{(2)}(v,n)$, $M(v/2,v/2,n)$, $M^*(v,n)$, $M^+(v,n)$ respectively, then, from (i), we have the lower bound

$$R(x,y) \ge S(u,\pi(y)) \tag{35}$$

and from (ii), the upper bound

$$R(x,(z,y]) \le S(\frac{\log x}{\log z},\ \pi(y)-\pi(z)) \tag{36}$$

In Proposition 2 we shall estimate each of the functions S and see that, for y and z as given,

$$S(u,\pi(y)),\ S(\frac{\log x}{\log z},\ \pi(y)-\pi(z)) = S(v,\pi(y))^{1+o(1)} \tag{37}$$

and that $\log\ S(v,\pi(y)) \gg_t\ u$.

From the results in Sections 1 and 2, we have estimated $R(x,z)$ in each case (as $\frac{z}{\log x} \to 0$ as $x \to \infty$) and seen that $R(x,z) = \exp(o(u))$.

Putting together this estimate of $R(x,z)$ with (34), (35), (36), (37) and the results in Proposition 2, we get the proofs of Theorems 2,3,4 and 5.

4. SOME COMBINATORICS

Proposition 1 For given positive integers v and n we have

(*i*) $$M(v,n) = \binom{v+n}{n}$$

(*ii*) $$M^{(2)}(v,n) = \sum_{j=0}^{v} \binom{n}{j}$$

(*iii*) If v is an even integer then $M(\frac{v}{2},\frac{v}{2},n) = \sum_{j=0}^{min(v/2,n)} \frac{n!(v/2+n-j)!}{j!^2(n-j)!^2(v/2-j)!}$

(*iv*) $$M^*(v,n) = \sum_{j=0}^{min(v,n)} \frac{(n+v-j)!}{j!(v-j)!(n-j)!}$$

(*v*) $$M^+(v,n) = \binom{v+2n}{2n}.$$

Proofs: We write "c. of X^u" to mean "coefficient of X^u".

(*i*) $$\begin{aligned} M(v,n) &= \sum_{u=0}^{v} \text{c. of } X^u \text{ in } (1+X+X^2+\ldots)^n \\ &= \text{c. of } X^v \text{ in } \frac{1}{(1-X)^{n+1}} = \binom{v+n}{n} \end{aligned}$$

(ii) $$M^{(2)}(v,n) = \sum_{u=0}^{v} \text{c. of } X^u \text{ in } (1+X)^n = \sum_{u=0}^{v} \binom{n}{u}$$

(iii) $$M(\frac{v}{2},\frac{v}{2},n) = \sum_{u=0}^{v/2}\sum_{w=0}^{v/2} \text{c. of } X^uY^w \text{ in } (1+X+Y+X^2+Y^2+\cdots)^n$$
$$= \text{c. of } X^{v/2}Y^{v/2} \text{ in } \frac{1}{(1-X)(1-Y)}\left(1+\frac{X}{1-X}+\frac{Y}{1-Y}\right)^n$$
$$= \sum_{i+j+k=n} \binom{n}{i,j,k}\binom{v/2}{j}\binom{v/2}{k}$$
$$= \sum_{j=0}^{n} \binom{n}{j}\binom{v/2}{j} \cdot \sum_{k=0}^{n-j} \binom{n-j}{k}\binom{v/2}{k}$$
$$= \sum_{j=0}^{n} \binom{n}{j}\binom{v/2}{j}\binom{v/2+n-j}{v/2}$$

(iv) $$M^*(v,n) = \sum_{u=0}^{v} \text{c. of } X^u \text{ in } (1+2X+2X^2+\cdots)^n$$
$$= \text{c. of } X^v \text{ in } \frac{(1+X)^n}{(1-X)^{n+1}}$$
$$= \sum_{j=0}^{min(v,n)} \frac{(n+v-j)!}{j!(v-j)!(n-j)!}$$

(v) $$M^+(v,n) = \sum_{u=0}^{v} \text{c. of } X^u \text{ in } (1+2X+3X^2+\cdots)^n$$
$$= \text{c. of } X^v \text{ in } \frac{1}{(1-X)^{2n+1}} = \binom{2n+v}{v}$$

We shall now estimate these functions in each case, taking $\tau \approx n/v$.

Proposition 2. Suppose that n and v are given positive integers such that $n = \tau v\{1+o(1)\}$ as $v \to \infty$. Then we have the uniform estimates

(i) $$\log\ M(v,n) \ = vF(\tau)\{1+o(1)\}$$

(ii) $$\log\ M^{(2)}(v,n) \ = vF^{(2)}(\tau)\{1+o(1)\}$$

(iii) $$\log\ M(\frac{v}{2},\frac{v}{2},n) \ = vF_2(\tau)\{1+o(1)\}$$

(iv) $$\log\ M^*(v,n) \ = vF_2(\tau)\{1+o(1)\}$$

(v) $$\log\ M^+(v,n) \ = vF(2\tau)\{1+o(1)\}$$

Proofs: (i) and (v) are immediate from the corresponding parts of Proposition 1.

In each of the remaining parts, the sums in Proposition 1 have at most n terms; and so by estimating the largest term in each sum we bring in an error of $O(\log\ n)$ which is acceptable.

(ii) If $v \geq n/2$ (i.e. $\tau \lesssim 2$) then the term $\binom{n}{n/2}$ is the largest, giving the value $\exp(\tau v \log\ 2\{1+o(1)\})$. If $v < n/2$ (i.e. $\tau \gtrsim 2$) then the term $\binom{n}{v}$ is the largest, giving rise to $\exp(vF^{(2)}(\tau)\{1+o(1)\})$.

(iii) and (iv) By comparing successive terms it is easy to show that in both cases the mazetamum term occurs where

$$\frac{j}{v} \approx \frac{1+t-\Delta}{2} \qquad \text{and} \qquad \Delta = (1+t^2)^{1/2}.$$

In both cases this gives rise to a term of size $\exp(v(\log(t+\Delta)+t\ \log(\frac{1+\Delta}{t}))\{1+o(1)\})$.

5. K-FREE INTEGERS.

Suppose k is a fixed integer, with $k \geq 2$, and define $\psi^{(k)}(x,y)$ to be the number of k-free integers in $S(x,y)$. In this section we show how difficult it is to estimate such functions when y is a fixed multiple of $\log x$.

If $y_0 < y \leq (\frac{1}{k-1} - \varepsilon)\log\ x$ then $\prod_{p \leq y} p^{k-1} < x$ by the prime number theorem and so

$$\psi^{(k)}(x,y) = k^{\pi(y)}.$$

Thus we have $\psi^{(k)}(x,y) = \psi(x,y)^{o(1)}$ in the range (23).

Of course it is well known that

$$\psi^{(k)}(x,x) \sim \psi(x,x)/\zeta(k)$$

and, by again imitating the methods of Ivič and Tenenbaum [12] it is possible to show that if $y > \log^{k/(k-1)+\varepsilon} x$ then $\psi^{(k)}(x,y) \sim \psi(x,y)/\zeta(k\alpha)$ where $\alpha = \alpha(x,y)$ is given by (13).

Actually, simply by noting the trivial inequality $\psi^{(2)}(x,y) \le \psi^{(k)}(x,y) \le \psi(x,y)$ we can derive, from Theorems 1 and 2, that $\psi^{(k)}(x,y) = \psi(x,y)^{1+o(1)}$ in the range (27).

So we are now only interested in the case where $y = t \log x$ for some fixed value of t. Let $p_1 < p_2 < \cdots < p_\pi$ be the primes up to y, and $u = \log x/\log y$.

Fix ε, $1 > \varepsilon > 0$. For any k-free integer $p_1^{a_1}\cdots p_\pi^{a_\pi}$ in $S(x,y)$ we have $a_1 + a_2 + \cdots + a_n \le u(1+\varepsilon)$ by the prime number theorem for any $x > x_\varepsilon$. Therefore

$$\psi^{(k)}(x, t \log x) \le \sum_{m \le u(1+\varepsilon)} \sum_{\substack{a_1+a_2+\cdots+a_\pi = m \\ \text{each } a_i \le k-1}} 1. \tag{38}$$

Also, if $a_1 + a_2 + \cdots + a_\pi \le u$ then $p_1^{a_1}\cdots p_\pi^{a_\pi} \le y^u = x$ and so

$$\psi^{(k)}(x, t \log x) \ge \sum_{m \le u} \sum_{\substack{a_1+a_2+\cdots+a_\pi = m \\ \text{each } a_i \le k-1}} 1. \tag{39}$$

In (38) and (39) we are summing over less than $2u$ values of m and so by estimating the largest term in each sum we bring in an error of $0(\log u)$ in the order of magnitude, which is negligible. Now $\sum_{\substack{a_1+a_2+\cdots+a_\pi=m \\ \text{each } a_i\le k-1}} 1$ is the coefficient of X^m in $(1 + X + X^2 + \cdots + X^{k-1})^\pi$, which gets larger as m gets closer to $\frac{\pi(k-1)}{2}$.

Therefore if $u \ge \frac{\pi(k-1)}{2}$ (i.e. $t \lesssim \frac{2}{k-1}$) then we can take $m = \frac{\pi(k-1)}{2}$ (to find the largest term) which gives

$$\log\ \psi^{(k)}(x, t \log x) = tu\ \log k\{1 + o(1)\}. \tag{40}$$

Henceforth we assume that $t \gtrsim \frac{2}{k-1}$ and so we take $m = u\{1+o(1)\}$ which gives

$$\log\ \psi^{(k)}(x, t \log x) = \log\ \lambda(k,\pi,u)\{1 + o(1)\} \tag{41}$$

where $\lambda(k,\pi,u)$, the coefficient of X^u in $(1 + X + \cdots + X^{k-1})^\pi$, equals

$$\sum_{\substack{n_1+2n_2+\cdots+(k-1)n_{k-1}=u \\ n_0+n_1+\cdots+n_{k-1}=\pi}} \binom{\pi}{n_0, n_1, \ldots, n_{k-1}}.$$

This has $\ll u^{k-1}$ terms, and so to determine the order of magnitude we again need only estimate the largest term in the sum.

Suppose that in the largest term we have $n_j = \alpha_j \pi$ for $j = 0, 1, 2, \ldots, k-1$, so that $\sum_{j=0}^{k-1} \alpha_j = 1$ and $\sum_{j=0}^{k-1} j\alpha_j = \frac{u}{\pi} \sim \frac{1}{t}$.

Then

$$\log \binom{\pi}{n_0, n_1, \ldots, n_{k-1}} = -\pi \sum_{j=0}^{k-1} \alpha_j \log \alpha_j \{1 + o(1)\}.$$

Thus the question that we have to answer is essentially:

$$minimize \quad \sum_{j=0}^{k-1} \alpha_j \log \alpha_j$$

subject to each $\alpha_j \geq 0$,

$$\sum_{j=0}^{k-1} \alpha_j = 1 \qquad \text{and} \qquad \sum_{j=0}^{k-1} j\alpha_j = \frac{1}{t}.$$

This is a classical problem in the theory of entropy which is notoriously difficult to solve in general.

6. REFERENCES

[1] N.G. de Bruijn, 'On the number of positive integers $\leq x$ and free of prime factors $> y$', *Nederl. Akad. Wetensch. Proc. Ser. A,* **54** (1951), 50-60.

[2] N.G. de Bruijn, 'On the number of positive integers $\leq x$ and free of prime factors $> y$, II', *Nederl. Akad. Wetensch. Proc. Ser. A,* **69** (1966), 239-247 = *Indag. Math.* **28** (1966), 239-247.

[3] K. Dickman, 'On the frequency of numbers containing prime factors of a certain relative magnitude', *Ark. Mat. Astr. Fys,* **22** (1930), 1-14.

[4] V. Ennola,'On numbers with small prime divisors', *Ann. Acad. Sci. Fenn. Ser. A. I.* **440** (1969), 16 pp.

[5] P. Erdös, 'Problem No. 136', *Wisk. Opgaven,* **21** (1963), 133-136.

[6] P. Erdös and J.H. van Lint, 'On the number of positive integers $\leq x$ and free of prime factors $> y$, *Simon Stevin,* 40 (1966-67), 73-76.

[7] A. Granville, 'On m-tuples of lattice points inside an n-dimensional tetrahedron', to appear.

[8] N.G. Gunderson, *Derivation of Criteria for the first Case of Fermat's Last Theorem and the Combination of these Criteria to produce a new lower bound for the exponent,* Ph.D. Thesis, Cornell University, (1948).

[9] A. Hildebrand, 'On the number of positive integers $\leq x$ and free of prime factors $> y$', *J. of Number Theory,* **22** (1986) 289-307.

[10] A. Hildebrand, 'On the number of prime factors of integers without large prime divisors', *J. of Number Theory,* **25** (1987) 81-106.

[11] A. Hildebrand and G. Tenenbaum, 'On integers free of large prime factors', *Trans. Amer. Math. Soc.*, **296** (1986), 265-290.

[12] A. Ivič and G. Tenenbaum, 'Local densities over integers free of large prime factors', *Quart. J. Math. Oxford,* **37** (1986), 401-417.

[13] B.V. Levin and A.S. Fanleib, 'Application of some integral equations to problems of number theory', *Uspehi Mat. Nauk.*, **22** (1967), 119-197 = *Russ. Math. Surveys* **22** (1967), 119-204.

[14] K.K. Norton, 'Numbers with small prime factors and the least k^{th} power non-residue', *Mem. Amer. Math. Soc.*, **106** (1971), 106 pp.

[15] R. Rankin, 'The difference between consecutive prime numbers', *J. London Math. Soc.*, **13** (1938), 242-247.

[16] S. Wigert, 'Sur l'ordre de grandeur du nombre des diviseurs d'un entier,' *Arkiv. für mat.*, **3**, (1907), 1-9.