Arch. Math., Vol. 43, 37–43 (1984)

0003-889X/84/4301-0037 $ 2.90/0

The distribution of the average prime divisor of an integer

By

JEAN-MARIE DE KONINCK and ALEKSANDAR IVIĆ

1. Introduction. Let us usual $\omega(n)$ and $\Omega(n)$ denote the number of distinct prime divisors and the total number of prime divisors of n respectively, and let $P(n)$ denote the largest prime factor of $n \geqq 2$. One may define the average prime factor of n as

$$P_*(n) = \frac{\beta(n)}{\omega(n)}, \quad \beta(n) = \sum_{p \mid n} p, \tag{1.1}$$

$$P^*(n) = \frac{B(n)}{\Omega(n)}, \quad B(n) = \sum_{p^\alpha \parallel n} \alpha p, \tag{1.2}$$

where as usual $p^\alpha \parallel n$ means that p^α divides n (p prime), but $p^{\alpha+1}$ does not. Thus one may consider $P_*(n)$ as the average of distinct prime factors of n, while $P^*(n)$ is the average of all prime factors of n, counted with their respective multiplicities. The functions $\beta(n)$ and $B(n)$ are additive, and they are "large" in the sense of Chapter 6 of [4]. Problems involving $\beta(n)$ and $B(n)$ have attracted much attention in recent years. Thus in [1] K. Alladi and P. Erdös showed

$$\sum_{2 \leqq n \leqq x} P(n) \sim \sum_{2 \leqq n \leqq x} \beta(n) \sim \sum_{2 \leqq n \leqq x} B(n) \sim \frac{\pi^2 x^2}{12 \log x}, \tag{1.3}$$

while P. Erdös and A. Ivić [6] showed that

$$\sum_{2 \leqq n \leqq x} \frac{\beta(n)}{P(n)} = x + O\left(\frac{x}{\log x}\right), \quad \sum_{2 \leqq n \leqq x} \frac{B(n)}{P(n)} = x + O\left(\frac{x}{\log x}\right), \tag{1.4}$$

and A. Ivić and C. Pomerance [10] proved a sharp asymptotic formula for the summatory function of $B(n)/\beta(n)$. Further results involving various estimates with $P(n)$, $\beta(n)$ and $B(n)$ are to be found in [3]–[10], and the purpose of this note is to sharpen (1.3) and to prove an asymptotic formula for the summatory function of $P_*(n)$, $P^*(n)$ and $P_*(n)/P^*(n)$. The results are contained in

Theorem 1. *For each fixed natural number m there exist computable constants $d_1 = \pi^2/12, d_2, \ldots, d_m$ such that*

$$\sum_{2 \leqq n \leqq x} P(n) = x^2 \left(\frac{d_1}{\log x} + \frac{d_2}{\log^2 x} + \ldots + \frac{d_m}{\log^m x} + O\left(\frac{1}{\log^{m+1} x}\right)\right), \tag{1.5}$$

and the formula remains true if $P(n)$ is replaced by $\beta(n)$ or $B(n)$.

Theorem 2. *For each fixed natural number m there exist computable constants* $e_1, e_2, \ldots, e_m, f_1, f_2, \ldots, f_m$ *such that* $0 < f_1 < e_1$ *and*

$$\sum_{2 \leqq n \leqq x} P_*(n) = x^2 \left(\frac{e_1}{\log x} + \frac{e_2}{\log^2 x} + \ldots + \frac{e_m}{\log^m x} + O\left(\frac{1}{\log^{m+1} x} \right) \right), \tag{1.6}$$

$$\sum_{2 \leqq n \leqq x} P^*(n) = x^2 \left(\frac{f_1}{\log x} + \frac{f_2}{\log^2 x} + \ldots + \frac{f_m}{\log^m x} + O\left(\frac{1}{\log^{m+1} x} \right) \right). \tag{1.7}$$

Theorem 3. *For each fixed natural number m there exist computable constants* $g_1, \ldots, g_m$ *such that*

$$\sum_{2 \leqq n \leqq x} \frac{P_*(n)}{P^*(n)} = x + x \left(\frac{g_1}{\log \log x} + \ldots + \frac{g_m}{(\log \log x)^m} + O\left(\frac{1}{(\log \log x)^{m+1}} \right) \right). \tag{1.8}$$

The notation used throughout is standard: p, q always denote primes, $a \mid b$ means that a divides b, $f(x) = O(g(x))$ and $f(x) \ll g(x)$ both mean $|f(x)| < C g(x)$ for $x \geqq x_0$ and some $C > 0$.

2. The necessary lemmas. In this section we present some lemmas which are necessary for the proofs of our theorems. These are

Lemma 1. *For each fixed natural number m there exist computable constants* $c_1 = 1/2$, $c_2, \ldots, c_m$ *such that*

$$\sum_{p \leqq x} p = x^2 \left(\frac{c_1}{\log x} + \frac{c_2}{\log^2 x} + \ldots + \frac{c_m}{\log^m x} + O\left(\frac{1}{\log^{m+1} x} \right) \right). \tag{2.1}$$

Lemma 2. *For each fixed pair of natural numbers m, i there exist computable constants* $c_{0,i} = \pi^2/6$, $c_{1,i}, \ldots, c_{m,i}$ *such that*

$$\sum_{n \leqq x} \frac{1}{n^2 \log^i \left(\frac{x}{n} + 1 \right)} = \frac{c_{0,i}}{\log^i x} + \frac{c_{1,i}}{\log^{i+1} x} + \ldots + \frac{c_{m,i}}{\log^{i+m} x} + O\left(\frac{1}{\log^{i+m+1} x} \right). \tag{2.2}$$

Lemma 3. *Let* $f(n) > 0$ *be an arithmetical function such that* $f(n) \ll \log n$. *Then for any fixed* $A > 0$

$$\sum_{2 \leqq n \leqq x} f(n)\, \beta(n) = \sum_{2 \leqq n \leqq x} f(n)\, P(n) + O(x^2 \log^{-A+1} x \log \log x), \tag{2.3}$$

$$\sum_{2 \leqq n \leqq x} f(n)\, B(n) = \sum_{2 \leqq n \leqq x} f(n)\, P(n) + O(x^2 \log^{-A+1} x \log \log x). \tag{2.4}$$

The proof of Lemma 1 is obtained by partial summation from the prime number theorem, while Lemma 2 follows by applying standard techniques from analysis. For Lemma 3 it is sufficient to prove (2.4) only, since the proof of (2.3) is quite similar. Write

$$\sum_{2 \leqq n \leqq x} f(n)\, B(n) = S_1 + S_2,$$

say, where in S_1 the summation is over $n \leqq x$ for which $P(n) \leqq n \log^{-A} n$, and in S_2 over the remaining n. Using

$$B(n) = \sum_{p^\alpha \| n} \alpha p \leqq P(n) \sum_{p^\alpha \| n} \alpha = P(n)\,\Omega(n)$$

and the elementary estimate

$$\sum_{n \leqq x} \Omega(n) \ll x \log\log x,$$

we obtain

$$S_1 \ll \sum_{P(n) \leqq n \log^{-A} n,\ n \leqq x} P(n)\,\Omega(n) \log x \ll x \log^{-A+1} x \sum_{n \leqq x} \Omega(n)$$
$$\ll x^2 \log^{-A+1} x \log\log x.$$

Consider now S_2 and suppose $q^\alpha \| n$, $q < P(n)$. Then

$$q^\alpha n \log^{-A} n < q^\alpha P(n) \leqq n,$$

which gives $\alpha q \leqq q^\alpha < \log^A n$. Therefore

$$S_2 = \sum_{2 \leqq n \leqq x} f(n)\,P(n) + O(x^2 \log^{-A+1} x \log\log x)$$
$$+ O\Big(\sum_{n \leqq x} \log n \sum_{q^\alpha \| n,\ q^\alpha < \log^A n} \alpha q\Big)$$
$$= \sum_{2 \leqq n \leqq x} f(n)\,P(n) + O(x^2 \log^{-A+1} x \log\log x) + O\Big(\log^{A+1} x \sum_{n \leqq x} \omega(n)\Big)$$
$$= \sum_{2 \leqq n \leqq x} f(n)\,P(n) + O(x^2 \log^{-A+1} x \log\log x).$$

3. Proof of theorems. First we present the proof of Theorem 1. Observe that by Lemma 3 with $f(n) = 1$ and $A = m + 3$ it is seen that (1.5) does indeed hold if $P(n)$ is replaced either by $\beta(n)$ or $B(n)$. To prove (1.5) write

$$\sum_{2 \leqq n \leqq x} P(n) = \sum_{pm \leqq x,\ P(m) \leqq p} p = \sum_{p \leqq x} p\psi(x/p, p), \tag{3.1}$$

where

$$\psi(x, y) = \sum_{n \leqq x,\ P(n) \leqq y} 1$$

is the function which represents the number of $n \leqq x$ for which $P(n) \leqq y$, and furthermore $\psi(x, y) = [x]$ if $y \geqq x$, while for $1 \leqq y < x$ various useful estimates for $\psi(x, y)$ exist in the literature (e.g. see [2]). But we have

$$\sum_{p \leqq x} p\psi(x/p, p) = \sum_{p \leqq \sqrt{x}} p\psi(x/p, p) + \sum_{\sqrt{x} < p \leqq x} p\psi(x/p, p) \tag{3.2}$$
$$= O(x^{3/2}) + \sum_{\sqrt{x} < p \leqq x} p[x/p] = \sum_{pn \leqq x} p + O(x^{3/2}),$$

since $\psi(x/p, p) = [x/p]$ for $p \geqq \sqrt{x}$. Observing that (2.1) remains true if $\log^i x$ ($i = 1, \ldots, m + 1$) is replaced by $\log^i (x + 1)$, we obtain then by using Lemma 2

$$\sum_{pn \leqq x} p = \sum_{n \leqq x} \sum_{p \leqq x/n} p = x^2 \left(\frac{d_1}{\log x} + \dots + \frac{d_m}{\log^m x} + O\left(\frac{1}{\log^{m+1} x} \right) \right), \tag{3.3}$$

and Theorem 1 follows now easily from (3.1), (3.2) and (3.3).

For the proof of Theorem 2 we remark that by Lemma 3 (with $A = m + 2$) it will be sufficient to prove the corresponding asymptotic expansions for

$$\sum_{2 \leqq n \leqq x} P(n)/\omega(n), \qquad \sum_{2 \leqq n \leqq x} P(n)/\Omega(n). \tag{3.4}$$

In both sums in (3.4) we may suppose that $P(n) \| n$, since

$$\sum_{2 \leqq n \leqq x,\, P^2(n) \mid n} P(n)/\omega(n) \leqq \sum_{p^2 m \leqq x} p \ll \sum_{p \leqq \sqrt{x}} p \left(\frac{x}{p^2} + 1 \right) \ll x \log\log x .$$

Also by the argument used in the proof of Lemma 3 we may suppose that $P(n) \in I$, where

$$I = [n \log^{-m-2} n, n].$$

Therefore if $\sum^*$ denotes summation over those n for which $P(n) \| n$ and $P(n) \in I$, then

$$\sum_{n \leqq x}^* \frac{P(n)}{\omega(n)} = \sum_k \frac{1}{k} \sum_{n \leqq x,\, \omega(n) = k}^* P(n) = \sum_k \frac{1}{k} S_k(x), \tag{3.5}$$

where k takes at most $O(\log x / \log\log x)$ values. Using Lemma 1 we have

$$S_1(x) = \sum_{p \leqq x} p + O(x^2 \log^{-m-2} x) = x^2 \left(\frac{c_1}{\log x} + \dots + \frac{c_m}{\log^m x} + O\left(\frac{1}{\log^{m+1} x} \right) \right),$$

while for $k \geqq 2$

$$S_k(x) = \sum_{n \leqq x,\, \omega(n) = k}^* P(n) = \sum_{\substack{p_1^{\alpha_1} \dots p_{k-1}^{\alpha_{k-1}} p \leqq x,\, p \in I \\ p_1 < \dots < p_{k-1} < p}} p \tag{3.6}$$

$$= \sum_{\substack{p_1^{\alpha_1} \dots p_{k-1}^{\alpha_{k-1}} \leqq x \\ p_1 < \dots < p_{k-1}}} \sum_{p_{k-1} < p \leqq x p_1^{-\alpha_1} \dots p_{k-1}^{-\alpha_{k-1}},\, p \in I} p .$$

To estimate the inner sum above we shall use Lemma 1. The contribution of those primes for which $p \leqq p_{k-1}$ is small, since $P(n) = p \in I$ implies $p_{k-1} \leqq \log^{m+2} n$, hence trivially

$$\sum_{\substack{p_1^{\alpha_1} \dots p_{k-1}^{\alpha_{k-1}} \leqq x \\ p_1 < \dots < p_{k-1} \leqq \log^{m+2} n}} \sum_{p \leqq p_{k-1}} p \ll \log^{2m+4} x \sum_{n \leqq x} 1 \ll x \log^{2m+4} x ,$$

and this estimate is uniform in k, so that summation over k in (3.5) will produce an error term which will certainly be $O(x^2 \log^{-m-1} x)$. Thus using Lemma 1 we see that the main contribution to $S_k(x)$ will be

$$x^2 \sum_{\substack{p_1^{\alpha_1} \dots p_{k-1}^{\alpha_{k-1}} \leqq x \\ p_1 < \dots < p_{k-1}}} p_1^{-2\alpha_1} \dots p_{k-1}^{-2\alpha_{k-1}} \sum_{i=1}^{m} c_i \log^{-i} \left(\frac{x}{p_1^{\alpha_1} \dots p_{k-1}^{\alpha_{k-1}}} + 1 \right), \tag{3.7}$$

and further we obtain

$$\sum_{\substack{p_1^{\alpha_1}\dots p_{k-1}^{\alpha_{k-1}}\leqq x\\ p_1<\dots<p_{k-1}}} p_1^{-2\alpha_1}\dots p_{k-1}^{-2\alpha_{k-1}}\log^{-i}\left(\frac{x}{p_1^{\alpha_1}\dots p_{k-1}^{\alpha_{k-1}}}+1\right) = \frac{g_0(i,k)}{\log^i x}+\frac{g_1(i,k)}{\log^{i+1}x}+\dots+\frac{g_m(i,k)}{\log^{i+m}x}+O\left(\frac{1}{\log^{i+m+1}x}\right) \tag{3.8}$$

for some suitable constants $g_0(i,k),\dots,g_m(i,k)$. In fact for some constant $h(i,j)$ ($j=0,1,\dots,m$) we have

$$g_j(i,k)=\sum_{d\leqq x,\ \omega(d)=k-1}\frac{h(i,j)\log^j d}{d^2}+O\left(\frac{\log^j x}{x}\right), \tag{3.9}$$

where the O-constant is uniform in k. From (3.5)–(3.9) we obtain then (1.6) with some computable constants $e_1, e_2, \dots, e_m$, since for each j

$$\sum_{k\geqq 2} k^{-1}g_j(i,k)=\sum_{d\leqq x}\frac{h(i,j)\log^j d}{(\omega(d)+1)\,d^2}+O\left(\frac{\log^j x\log\log x}{x}\right)$$
$$=h(i,j)\sum_{d=1}^{\infty}\frac{\log^j d}{(\omega(d)+1)\,d^2}+O\left(\frac{\log^j x\log\log x}{x}\right),$$

and by collecting various terms of the form $x^2\log^{-t}x$ ($t=1,2,\dots,m$) we obtain (1.6).

The proof of (1.7) is completely analogous to the proof of (1.6) and this therefore omitted, but note that we obtain $0<f_1<e_1$, which means that the average order of $P_*(n)$ is larger than the average order of $P^*(n)$.

As for Theorem 3, it may be noted that using the method of proof of Lemma 3 it is sufficient to estimate

$$\sum_{n\leqq x}{}^{*} P_*(n)/P^*(n)=\sum_{n\leqq x}{}^{*}\frac{\Omega(n)\,P(n)}{\omega(n)\,B(n)}+O(x\log^{-A}x), \tag{3.10}$$

since

$$\sum_{n\leqq x} 1/B(n)=x\exp\{-(2\log x\log\log x)^{1/2}+O((\log x\log\log\log x)^{1/2})\}, \tag{3.11}$$

as proved in [9]. Here $\sum^*$ denotes summation over $n\leqq x$ such that $P(n)\,\|\,n$ and $P(n)\in I=[n\log^{-A}n, n]$, where $A>0$ is a sufficiently large constant. The condition $P(n)\in I$ implies $q\leqq\log^A n$ if $q^\alpha\,\|\,n$, $q<P(n)$, so that in $\sum^*$ in (3.10) one may also replace $B(n)$ by $P(n)$ with a manageable error, since using (3.11) we have

$$\sum_{n\leqq x}{}^{*}\frac{\Omega(n)}{\omega(n)}-\sum_{n\leqq x}{}^{*}\frac{\Omega(n)\,P(n)}{\omega(n)\,B(n)}$$
$$\ll\log x\sum_{2\leqq n\leqq x}\frac{1}{B(n)}\sum_{q^\alpha\,\|\,n,\ q<P(n),\ q\leqq\log^A n}\alpha q$$
$$\ll\log^{A+2}x\sum_{2\leqq n\leqq x}\frac{1}{B(n)}\ll x\log^{-A}x.$$

Thus all we are left with is the sum of quotients of $\Omega(n)$ and $\omega(n)$, i.e.

$$\sum_{n\leq x}{}^{*}\,\Omega(n)/\omega(n) = \sum_{2\leq n\leq x}\Omega(n)/\omega(n) + O(x\log^{1-A}x)$$
$$= x + x\left(\frac{g_1}{\log\log x} + \ldots + \frac{g_m}{(\log\log x)^m} + O\left(\frac{1}{(\log\log x)^{m+1}}\right)\right),$$

if $A > 1$, when one used the asymptotic formula (4.3) of [4]. This proves Theorem 3.

4. Remarks. Instead of summing $P_*(n)$ or $P^*(n)$ one may sum their reciprocals, and this was investigated in [7] and [8]. In that case one has, following the proof of (1.7) in [8],

$$\sum_{2\leq n\leq x}\frac{1}{P_*(n)} = \sum_{2\leq n\leq x}\frac{\omega(n)}{\beta(n)} = (\sqrt{2}+o(1))\left(\frac{\log x}{\log\log x}\right)^{1/2}\sum_{2\leq n\leq x}\frac{1}{\beta(n)},$$
$$\sum_{2\leq n\leq x}\frac{1}{P^*(n)} = \sum_{2\leq n\leq x}\frac{\Omega(n)}{B(n)} = (\sqrt{2}+o(1))\left(\frac{\log x}{\log\log x}\right)^{1/2}\sum_{2\leq n\leq x}\frac{1}{B(n)},$$

while

$$\sum_{2\leq n\leq x}\frac{1}{\beta(n)} = x\exp\{-(2\log x\log\log x)^{1/2} + O((\log x\log\log\log x)^{1/2})\}.$$

The last formula was proved in [9] and sharpened in [10]; the result remains true if $\beta(n)$ is replaced by $P(n)$ or $B(n)$ (see (3.11)).

We may also remark that (1.8) (with possibly different g_i's) remains true if the left-hand side is replaced by $\sum_{2\leq n\leq x} P^*(n)/P_*(n)$. This follows from the fact that an asymptotic formula for $\sum_{2\leq n\leq x}\omega(n)/\Omega(n)$, analogous to (4.3) of [4] also holds, and may be established by the same method of proof. Also the function on the right-hand side of (1.8) could be replaced by a more precise expression if instead of (4.3) of [4] one uses the more refined estimate for $\sum_{2\leq n\leq x}\Omega(n)/\omega(n)$, proved in chapter 5 of [4].

The method of proof of Theorem 2 could be also used to yield estimates of the type (1.6) and (1.7) for the more general sums $\sum_{2\leq n\leq x} f(n)P(n)$, where $f(n)$ ($\gg 1$) is an additive, prime-independent function of moderate growth.

References

[1] K. ALLADI and P. ERDÖS, On an additive arithmetic function. Pacific J. Math. **71**, 275–294 (1977).

[2] N. G. DE BRUIJN, On the number of positive integers $\leqq x$ and free of prime factors $> y$. Indag. Math. **13**, 50–60 (1951) and II ibid. **28**, 239–247 (1966).

[3] J.-M. DE KONINCK, P. ERDÖS and A. IVIĆ, Reciprocals of large additive functions. Canadian Math. Bull. **24**, 225–231 (1981).

[4] J.-M. DE KONINCK and A. IVIĆ, Topics in arithmetical functions. Notás de Matemática **72**, Amsterdam 1980.

[5] J.-M. DE KONINCK et A. IVIĆ, Sommes de réciproques de grandes fonctions additives. Publs. Inst. Mathém. Belgrade **35** (1984), in print.

[6] P. ERDÖS and A. IVIĆ, Estimates for sums involving the largest prime factor of an integer and certain related additive functions. Studia Scien. Math. Hungarica **15**, 183–199 (1980).

[7] P. Erdös and A. Ivić, On sums involving reciprocals of certain arithmetical functions. Publs. Inst. Mathém. Belgrade **32**, 49–56 (1982).
[8] P. Erdös, A. Ivić and C. Pomerance, On sums involving reciprocals of the largest prime factor of an integer. To appear.
[9] A. Ivić, Sums of reciprocals of the largest prime factor of an integer. Arch. Math. **36**, 57–61 (1980).
[10] A. Ivić and C. Pomerance, Estimates of certain sums involving the largest prime factor of an integer. Coll. Math. Soc. J. Bolyai **34**, Topics in classical number theory, Amsterdam 1984.

Eingegangen am 30. 9. 1983

Anschrift der Autoren:

Jean-Marie De Koninck
Départment de Mathématiques
Université Laval
Québec, G1K 7P4
Canada

Aleksandar Ivić
Katedra Matematike
RGF-a Universiteta u Beogradu
Djušina 7, 11000 Beograd
Yugoslavia