ON THE LARGEST PRIME DIVISORS OF AN INTEGER

Jean-Marie De Koninck (Université Laval)

1. Introduction

Several results of number theory can be expressed in probabilistic terms and, for others, the simplest proof is by probabilistic methods. Simply take the uniform distribution on the consecutive integers 1, 2, ..., N. Then arithmetic functions, when restricted to the integers 1 through N, become random variables and arithmetic means are expectations. The power of probabilistic methods lies in the fact that divisibility by distinct primes are almost independent events. On the other hand, most problems remain challenging since the errors generated by the not exact independence can be dominating in a problem when one faces an increasing number of primes. The best example is the study of large prime divisors where the results do not resemble those which one would get for independent random variables.

Our work is following the number theoretic tradition, but in the light of the above probability space one can translate most statements into ones on random variables and expectations.

Given an integer $n > 1$, let $p(n)$ be the smallest prime factor of n and $P(n)$ $(= P_1(n))$ be the largest prime factor of n. More generally, for each integer $k \geq 1$ and for each positive integer n such that $\Omega(n) \geq k$ (here $\Omega(n) = \sum_{p^\alpha \| n} \alpha$), let $P_k(n)$ be the k^{th} largest prime factor of n. Hence, for an integer $n > 1$ such that $\Omega(n) = \ell$, we have

$$p(n) = P_\ell(n) \leq P_{\ell-1}(n) \leq \ldots \leq P_2(n) \leq P_1(n) = P(n).$$

We shall write $P(1) = 1$.

J. Galambos et al. (eds.), Extreme Value Theory and Applications, 447–462.

It is also common to use the notation $p_k(n)$ to denote the k^{th} distinct prime factor of n, so that

$$p(n) = p_1(n) < p_2(n) < \ldots < p_{\omega(n)} = P(n),$$

where $\omega(n) = \sum_{p|n} 1$ is the number of distinct prime factors of n.

For large values of k, the behaviour of the function $p_k(n)$ has been established by Galambos [9], who proved the following two results:

(i) Given $\varepsilon > 0$, if $k = k(N) \to +\infty$ in such a way that $k(N) \leq \log\log N - (\log\log N)^{\frac{1}{2}+\varepsilon}$, then

$$\lim_{N\to\infty} \frac{1}{N}\#\{n \leq N : \log\log p_k(n) \leq k + z\sqrt{k}\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-\frac{1}{2}t^2}\,dt.$$

(ii) Given $\varepsilon > 0$, if $k = k(N) \to +\infty$ in such a way that $k(N) \leq (1-\varepsilon)\log\log N$, then, if $z > 0$,

$$\lim_{N\to\infty} \frac{1}{N}\#\{n \leq N : \log\log p_{k+1}(n) - \log\log p_k(n) \leq z\} = 1 - e^{-z}.$$

These results of J. Galambos were in 1976 a turning point in probabilistic number theory and have since given way to a great variety of results geared towards a better understanding of the prime factorization of integers.

The above statements (i) and (ii) relate to the so-called "intermediate prime divisors" of integers; they were recently generalized by De Koninck and Galambos [4].

On the other hand, the distribution of the small prime divisors of an integer n, that is of $p_k(n)$ for a fixed $k \in \mathbb{N}$, can be handled using the Prime Number Theorem and classical results on asymptotic estimates of sums of arithmetical functions.

The study of the distribution of the large prime divisors of an integer n, that is that of $P_k(n)$ for a fixed $k \in \mathbb{N}$ or of $P(n,Q)$, the largest prime factor of n belonging to a particular set of primes Q, is generally recognized as being a more difficult problem (than that of $p_k(n)$ for fixed $k \in \mathbb{N}$) and is the object of this paper.

First we present a survey of results concerning the behaviour of $P_k(n)$, $k \geq 1$, and of $P(n,Q)$. Secondly, given a fixed integer $k \geq 1$ and a large number x, we establish the most frequent values taken by $P_k(n)$ among the integers $n \leq x$. We also establish the most frequent value of $P(n,Q)$, when Q is a set of primes of positive density $\delta < 1$ satisfying a certain regularity condition.

2. The functions $P(n)$, $P_k(n)$ and $P(n,Q)$

We start with the function $P(n)$ which has mostly been studied through the function

$$\Psi_1(x,y) = \Psi(x,y) \stackrel{\text{def}}{=} \#\{n \leq x : P(n) \leq y\}.$$

The best estimate concerning the asymptotic behaviour of $\Psi(x,y)$ is that obtained by Hildebrand [13] in 1986:

$$\Psi(x,y) = x\rho(u)\left(1 + O\left(\frac{\log(u+1)}{\log y}\right)\right), \tag{2.1}$$

where $u = \log x/\log y$, uniformly for each $\varepsilon > 0$ in the range

$$\exp\left\{(\log\log x)^{\frac{5}{3}+\varepsilon}\right\} \leq y \leq x,$$

$\rho(u)$ being the continuous solution of $u\rho'(u) = -\rho(u-1)$ with initial condition $\rho(u) = 1$ for $0 \leq u \leq 1$. A more detailed study of $\Psi(x,y)$ can be found in the recent book of G. Tenenbaum [18].

The function $P_2(n)$ has also been extensively studied (see the book of H. Riesel [17]). It appears in several number theory problems, in particular the factorization of large numbers (see Hafner and McCurley [12] and Knuth and Pardo [16]).

The study of the function $P(n)$, mainly through the function $\Psi(x,y)$, is quite complex. This is mainly due to the fact that $P(n)$ is very erratic. The mean behaviour of the function $P_k(n)$ has been studied by Alladi and Erdős [1] who obtained that

$$\sum_{n\leq x} P_k(n) = A_k\frac{x^{1+\frac{1}{k}}}{\log^k x} + O\left(x^{1+\frac{1}{k}}\frac{\log\log x}{\log^{k+1} x}\right) \qquad (A_k > 0) \tag{2.2}$$

for $k \geq 1$ fixed. This formula was improved for $k = 1$ by De Koninck and Ivić [5] (with $A_1 = \frac{\pi^2}{12}$) and for $k \geq 2$ by Ivić [14]. Although the mean behaviour of $P(n)$ is easy to obtain, that of the function $1/P(n)$ is very difficult to grasp; in 1986, Erdős, Ivić and Pomerance [8] nevertheless proved that

$$\sum_{n\le x}\frac{1}{P(n)}=(1+o(1))\frac{x}{L(x)}, \tag{2.3}$$

where

$$L(x)^{-1}=\int_2^x \rho\left(\frac{\log x}{\log t}\right)t^{-2}\,dt. \tag{2.4}$$

Using the estimate

$$\rho(u)=\exp\left\{-u(\log u+\log\log u-1+o(1))\right\}, \tag{2.5}$$

one can prove that $L(x)$ is a *slowly oscillating* function (i.e. such that $\lim_{x\to\infty}\frac{L(cx)}{L(x)}=1$ for each constant $c>0$) and that

$$L(x)=\exp\left\{(1+o(1))\sqrt{2\log x\log\log x}\right\}. \tag{2.6}$$

On the other hand, the study of the behaviour of $P_2(n)$ rests primarily on the function

$$\Psi_2(x,y)\stackrel{\text{def}}{=}\#\{n\le x: P_2(n)\le y\}$$

(see Riesel [17]).

The best estimate concerning the asymptotic development of $\Psi_2(x,y)$ is certainly that of Hafner and McCurley [12]:

$$\Psi_2(x,y)=x\rho_2(u)\left(1+O\left(\frac{1}{\log y}\right)\right),$$

valid uniformly for $2\le y\le x$, where $\rho_2(u)$ is the continuous solution of the equation

$$u\rho_2'(u)+\rho_2(u-1)=\rho(u-1)$$

with initial condition $\rho_2(u)=1$ for $0\le u\le 1$. While, as confirmed by the relation (2.5), the function $\rho(u)$ decreases very rapidly to 0 as $u\to\infty$, the function $\rho_2(u)$ tends to 0 much more slowly since it satisfies

$$\rho_2(u)=\frac{e^\gamma}{u}\left(1+O\left(\frac{1}{u}\right)\right),\qquad (u\ge 1)$$

(see Knuth and Pardo [16]). More generally, one can prove that

$$\Psi_k(x,y)\stackrel{\text{def}}{=}\#\{n\le x: P_k(n)\le y\}=x\rho_k(u)\left(1+O\left(\frac{1}{\log y}\right)\right),$$

uniformly for $2\le y\le x$, where $\rho_k(u)$ is the continuous solution of $u\rho_k'(u)+\rho_k(u-1)=\rho_{k-1}(u-1)$, with initial condition $\rho_k(u)=1$ for $0\le u\le 1$.

In De Koninck [3], we obtained that, given an integer $A\ge 1$, there exist constants $\lambda_2^{(1)},\lambda_2^{(2)},\ldots,\lambda_2^{(A)}$ such that

$$\sum_{\substack{4\le n\le x\\ \Omega(n)\ge 2}}\frac{1}{P_2(n)}=\lambda_2^{(1)}\frac{x}{\log x}+\lambda_2^{(2)}\frac{x}{\log^2 x}+\ldots+\lambda_2^{(A)}\frac{x}{\log^A x}+O\left(\frac{x}{\log^{A+1}x}\right),$$

where $\lambda_2^{(1)}=\lambda_2=\sum_{m=1}^{\infty}\frac{1}{m}\sum_{p\ge P(m)}\frac{1}{p^2}=\sum_p\frac{1}{p^2}\prod_{q\le p}\left(1-\frac{1}{q}\right)^{-1}=1.254\ldots$, thereby improving an earlier result of Erdős and Ivić [7]. More generally, in [3] we proved that, given an integer $k\ge 2$,

$$\sum_{\substack{2^k\le n\le x\\ \Omega(n)\ge k}}\frac{1}{P_k(n)}=\lambda_k\frac{x(\log\log x)^{k-2}}{\log x}\left(1+O\left(\frac{1}{\log\log x}\right)\right),$$

where $\lambda_k=\frac{1}{(k-2)!}\lambda_2$. Note the drastic difference in behaviour between $\sum 1/P(n)$ and $\sum 1/P_k(n)$ (for $k\ge 2$).

What if one is interested in the largest prime factor of an integer n amongst the primes belonging to a congruence class, say those primes $p\equiv 1 \pmod 4$? More generally, given a set Q of prime numbers of positive density $\delta<1$ satisfying the regularity condition

$$\pi(x,Q)\stackrel{\text{def}}{=}\sum_{p\le x,\ p\in Q}1=\delta\,\mathrm{Li}(x)+O\left(\frac{x}{\log^B x}\right), \tag{2.7}$$

where $\mathrm{Li}(x)=\int_2^x\frac{dt}{\log t}$ and $B>2$ is a constant, and let $P(n,Q)=1$ if no prime factor of n belongs to Q and otherwise $P(n,Q)=\max\{p: p|n\wedge p\in Q\}$, what can one say about the behaviour of the function $P(n,Q)$. Let $(n,Q)=1$ if all prime divisors of n do not belong to Q and otherwise let (n,Q) be the greatest divisor of n whose prime factors belong to Q. One can show that

$$\sum_{\substack{n\le x\\(n,Q)>1}} P(n,Q)=(1+o(1))c(Q)\frac{x^2}{\log x},\tag{2.8}$$

for a constant $c(Q)>0$ (see Ivić [15]) the main contribution to the sum coming from those integers n which are prime numbers belonging to Q.

In De Koninck [3], we obtained that if Q is a set of prime numbers of density $0<\delta<1$ satisfying the condition (2.7), then there exists a positive constant $\eta(Q)$ such that

$$\sum_{\substack{n\le x\\(n,Q)>1}} \frac{1}{P(n,Q)}=\left(1+O\left(\frac{1}{\log\log x}\right)\right)\eta(Q)\frac{x}{(\log x)^\delta},\tag{2.9}$$

an estimate which is much less complex than the one provided by the relations (2.3), (2.4) and (2.6) in the case where Q is the set of all primes. Hence the irregularity of $\sum_{n\le x}1/P(n)$ is considerably diminished if one ignores a finite proportion of the primes. The proof of (2.9) is based on results obtained by Goldston and McCurley (see [10] and [11]), namely the fact that

$$\Psi(x,y,Q)\stackrel{\text{def}}{=}\sum_{\substack{n\le x\\P(n,Q)\le y}}1=x\rho_\delta(u)\left(1+O\left(\frac{1}{\log y}\right)\right)\tag{2.10}$$

uniformly for $u\ge1$ and $y\ge 3/2$, where $\rho_\delta(u)$ is the continuous solution of

$$u\rho_\delta'(u)=-\delta\rho_\delta(u-1)\qquad(u\ge1),$$

with initial condition $\rho_\delta(u)=1$ for $0\le u\le1$, and the fact that, as $u\to\infty$,

$$\rho_\delta(u)=\frac{e^{\gamma\delta}}{\Gamma(1-\delta)}\frac{1}{u^\delta}\left(1+O\left(\frac1u\right)\right).\tag{2.11}$$

Set

$$P_k(n,Q)\stackrel{\text{def}}{=}\begin{cases}P\left(\frac{n}{P_1(n,Q)\dots P_{k-1}(n,Q)},Q\right) & \text{if }\Omega(n,Q)\ge k,\\ 0 & \text{otherwise,}\end{cases}$$

if $k\ge2$, where $P_1(n,Q)\equiv P(n,Q)$ and $\Omega(n,Q)=\sum_{p^\alpha\|n,\ p\in Q}\alpha$. In De Koninck and Ivić [6], various estimates related to the behaviour of $P(n,Q)$ and $P_k(n,Q)$ are established. In particular, it is shown that, if $k\ge2$,

$$\sum_{\substack{n\le x\\(n,Q)>1}}\frac{P_k(n,Q)}{P(n,Q)}=(1+o(1))C_k(Q)\frac{x}{(\log x)^\delta},$$

for some positive constant $C_k(Q)$. Concerning this last estimate, one observes that all the sums of $P_k(n,Q)/P(n,Q)$, for any $k\ge2$, are of the same order, namely $x/(\log x)^\delta$ which does not depend on k, while it was proven by Erdős and Alladi [2] that, for fixed $k\ge2$,

$$\sum_{2\le n\le x}\frac{P_k(n)}{P(n)}=(1+o(1))a_k\frac{x}{(\log x)^{k-1}},$$

for some positive constant a_k, thereby establishing that the order of the sums of $P_k(n)/P(n)$ changes with k.

In De Koninck [3], we proved that if Q satisfies (2.7), then the median value of $P(n,Q)$ for the integers $n\le x$ is $n^{\kappa+o(1)}$, where κ is the unique solution of

$$\rho_\delta\left(\frac1\kappa\right)=\frac12.\tag{2.12}$$

Here, to look for the median value κ of $P(n,Q)$ amounts to seek, among the numbers $\kappa\in]0,1[$, the one for which

$$\sum_{\substack{x^\kappa<p\le x\\p\in Q}}\ \sum_{\substack{n\le x\\P(n,Q)=p}}1\sim\frac x2.$$

The proof uses estimates (2.10) and (2.11).

It is interesting to note that if the density δ of the set Q satisfies $\delta>\dfrac{1}{2\log2}$, one can show that the median value of $P(n,Q)$ is $n^{\kappa+o(1)}$, where $\kappa=e^{-\frac{1}{2\delta}}$, while for small values of δ, one can obtain, using (2.10) and (2.11), that the corresponding value of κ becomes very little and namely satisfies $\kappa\sim\frac{1}{2^{1/\delta}}$ if $\delta\to0$. In all cases, note that the median value of $P(n,Q)$ is considerably smaller than its mean value which, according to (2.8), is $n^{1+o(1)}$.

It is also shown in De Koninck [3] that, for each integer $k\ge1$, the median value of $P_k(n)$ for the integers $n\le x$ is $n^{\kappa+o(1)}$, where $\kappa=\kappa(k)$ is the unique solution of the equation

$$\rho_k\left(\frac1\kappa\right)=\frac12.$$

Since $\rho(u)=1-\log u$ for $1\le u\le2$, one has $\frac12=\rho\left(\frac1\kappa\right)=1+\log\kappa$, hence $\kappa=\kappa(1)=\frac{1}{\sqrt e}=0.606\ldots$, a result already obtained by Wunderlich et

Selfridge [19]. Using the tables giving the values of $\rho_2(u)$ (see H. Riesel [17]), one obtains $\kappa = \kappa(2) = 0.21\ldots$ thus improving the estimate $\kappa = 0.24$ obtained empirically by Wunderlich and Selfridge [19]. Note that we have thus established that the median value of $P_2(n)$ is $n^{0.21\ldots+o(1)}$, a value which is "not so far" from the mean value of $P_2(n)$, which can be seen by (2.2) to be $n^{0.5+o(1)}$, a definite contrast with the case of $P(n,Q)$.

3. On most frequent values of $P(n)$, $P_k(n)$ and $P(n,Q)$

Theorem 1. *Let $f(p) = f_1(p,x) \stackrel{\text{def}}{=} \#\{n \le x : P(n) = p\}$. For large x, the maximum value of $f(p)$ is attained at primes p satisfying*

$$p = e^{\sqrt{\frac{1}{2}\log x \log\log x}\left(1+\frac{\lambda(x)}{2}+o\left(\frac{1}{\log\log x}\right)\right)}, \tag{3.1}$$

where

$$\lambda(x) \stackrel{\text{def}}{=} \frac{\log\log\log x}{\log\log x},$$

in which case

$$f(p) = xe^{-\sqrt{2\log x\log\log x}\left(1+\frac{\lambda(x)}{2}-\frac{1}{2}\frac{2+\log 2+o(1)}{\log\log x}\right)}. \tag{3.2}$$

Remark. Again it is interesting to note that these results add to the "wild behaviour" of the function $P(n)$. Indeed, on the one hand, as we saw earlier, both the mean value and the median value of $P(n)$ are each in the neighbourhood of a positive power of n, while on the other hand the most frequent value of $P(n)$ is, according to (3.1), smaller than any positive power of n.

Proof. Assume that x is large. To start we show that, if $0 < a < b$ and if p is a prime such that

$$e^{\sqrt{a\log x\log\log x}} < p < e^{\sqrt{b\log x\log\log x}}, \tag{3.3}$$

then

$$f(p) > \frac{x}{e^{c\sqrt{\log x\log\log x}(1+O(\lambda(x)))}}, \tag{3.4}$$

where

$$c = \sqrt{b} + \frac{1}{2\sqrt{a}}.$$

Indeed let $u = \frac{\log x}{\log p}$ and p in the range (3.3), then clearly

$$u < \frac{1}{\sqrt{a}}\sqrt{\frac{\log x}{\log\log x}} \quad \text{and} \quad \log u < \frac{1}{2}\log\log x. \tag{3.5}$$

Using this, the fact that $f(p) = \Psi(\frac{x}{p}, p)$ and (2.1), we have, in the range (3.3),

(3.6)
$$f(p) = \frac{x}{p}\rho(u-1)\left(1+O\left(\frac{\log u}{\log p}\right)\right) = \frac{x}{p}\rho(u-1)\left(1+O\left(\sqrt{\frac{\log\log x}{\log x}}\right)\right).$$

Let $\xi = \xi(u)$ be the unique solution of $e^\xi = 1 + u\xi$. Formulas (47) and (61) in Tenenbaum [18] (pp. 412 and 417) give, for large x,

$$\frac{\rho(u-1)}{\rho(u)} = e^{\xi(u)}(1+O(\tfrac{1}{u})) = (1+o(1))u\log u. \tag{3.7}$$

Hence, if x is large enough, combining (3.3), (3.5), (3.6), (3.7) and (2.5), we have

$$\begin{aligned} f(p) &> \frac{x}{e^{\sqrt{b\log x\log\log x}+u\log u+u\log\log u}} \\ &> \frac{x}{e^{\sqrt{b\log x\log\log x}+\frac{1}{2\sqrt{a}}\sqrt{\log x\log\log x}(1+2\lambda(x))}}, \end{aligned}$$

since, in the range (3.3), $u\log u < \dfrac{\sqrt{\log x\log\log x}}{2\sqrt{a}}$ and

$$u\log\log u < \frac{1}{\sqrt{a}}\frac{\sqrt{\log x}}{\sqrt{\log\log x}}\log\log\log x = \frac{\lambda(x)}{\sqrt{a}}\sqrt{\log x\log\log x}.$$

We have thus established (3.4) with $c = \sqrt{b} + \frac{1}{2\sqrt{a}}$.

We will now show that the maximum of $f(p)$ must be attained within the range (3.3) for some constants $0 < a < b$. Indeed, using the elementary estimate $\Psi(x,y) \ll xe^{-\frac{1}{2}\frac{\log x}{\log y}}$ (see Tenenbaum [18]), we clearly have that, for $p < \exp\{(\log x)^{1/4}\}$, $f(p) = \Psi(\frac{x}{p}, p) \ll x\exp\{-\frac{1}{2}\frac{\log x}{\log p}\} \ll$

$x\exp\{-\frac{1}{2}(\log x)^{3/4}\}$. On the other hand, in the range $\exp\{(\log x)^{1/4}\} \le p < \exp\{\frac{1}{M(x)}\sqrt{\log x \log\log x}\}$ (with $M(x) \to \infty$), we may use the estimates (2.1) and (2.5) and deduct that

$$f(p) = \Psi(\frac{x}{p}, p) \ll \frac{x}{p} e^{-u\log u} < x e^{-u\log u} \ll \frac{x}{e^{\frac{M(x)}{2}\sqrt{\log x \log\log x}}}.$$

On the other hand, using the trivial estimate $\Psi(\frac{x}{p}, p) \le \frac{x}{p}$, it is clear that if $M(x)$ tends to infinity with x and if $p > \exp\{M(x)\sqrt{\log x \log\log x}\}$, then

$$f(p) = \Psi(\frac{x}{p}, p) \ll \frac{x}{e^{M(x)\sqrt{\log x \log\log x}}}.$$

From these estimates, we conclude that the maximum value of $f(p)$ must be attained in the range (3.3) for some positive values of a and b.

We now show that with

$$p = e^{\sqrt{a\log x \log\log x}(1+o(1))}, \tag{3.8}$$

where a is any positive number, then

$$f(p) = x e^{-c\sqrt{\log x \log\log x}(1+o(1))}, \quad \text{with} \quad c = \sqrt{a} + \frac{1}{2\sqrt{a}}. \tag{3.9}$$

First, if we fix a prime p satisfying (3.8), define $\xi(x)$ implicitly by the relation

$$p = e^{\sqrt{a\log x \log\log x}(1+\xi(x))}.$$

Then, with $u = \frac{\log x}{\log p}$, it is clear that

$$u = \frac{1}{\sqrt{a}} \frac{\sqrt{\log x}}{\sqrt{\log\log x}(1+\xi(x))}$$

and thus successively

$$\begin{aligned} \log u &= \frac{1}{2}\log\log x - \frac{1}{2}\log\log\log x - \frac{1}{2}\log a + o(1) \\ \log\log u &= \log\log\log x - \log 2 + o(1). \end{aligned}$$

From these estimates, it follows that

$$\log u + \log\log u - 1 = \frac{1}{2}\log\log x + \frac{1}{2}\log\log\log x - \left(\frac{1}{2}\log a + \log 2 + 1\right) + o(1)$$

and therefore that, recalling (2.5) and in view of (3.7),

$$\begin{aligned} \log\rho(u-1) &= \log\rho(u)\left(1 + O(\frac{1}{u})\right) \\ &= -u(\log u + \log\log u - 1 + o(1))\left(1 + O(\frac{1}{u})\right) \\ &= -u(\log u + \log\log u - 1 + o(1)) \\ &= -\frac{1}{2\sqrt{a}} \frac{\sqrt{\log x \log\log x}}{1+\xi(x)} \left(1 + \lambda(x) - \frac{\log a + 2\log 2 + 2 + o(1)}{\log\log x}\right). \end{aligned}$$

Then, since $f(p) = \Psi(\frac{x}{p}, p)$, using (2.1), we have

$$\begin{aligned} f(p) &= \frac{x}{p}\rho(u-1)\left(1 + O\left(\frac{\log u}{\log p}\right)\right) \\ &= \frac{x}{p}\rho(u-1)\left(1 + O\left(\sqrt{\frac{\log\log x}{\log x}}\right)\right) \\ &= \frac{x}{p} e^{\log\rho(u-1)}(1+o(1)), \end{aligned}$$

from which it follows that

(3.10)

$$f(p) = \frac{x}{e^{\sqrt{a\log x\log\log x}(1+\xi(x))} \cdot e^{\frac{1}{2\sqrt{a}}\sqrt{\log x \log\log x}\left(1+\lambda(x) - \frac{\log a + 2\log 2 + 2 + o(1)}{\log\log x}\right)/(1+\xi(x))}}.$$

Since both $\lambda(x)$ and $\xi(x)$ are $o(1)$, one easily obtains that

$$f(p) = \frac{x}{e^{\sqrt{\log x \log\log x}\left\{\left(\sqrt{a} + \frac{1}{2\sqrt{a}}\right)(1+o(1))\right\}}},$$

which implies (3.9).

Now using (3.9) it is clear that the maximum value of $f(p)$ in the range (3.3) is attained when $c = \sqrt{2}$ with a corresponding value of a equal to $\frac{1}{2}$, since the minimum value of $\sqrt{a} + \frac{1}{2\sqrt{a}}$, for $a > 0$, is equal to $\sqrt{2}$ and is attained when $a = \frac{1}{2}$.

To complete the proof of (3.1) and (3.2), note that, using (3.10) with $a = \frac{1}{2}$, it follows that

$$f(p) = \frac{x}{e^{\sqrt{\frac{1}{2}\log x \log\log x}\{\Delta(x)\}}},$$

where

$$\Delta(x) = 2+\lambda(x)+\xi^2(x)-\lambda(x)\xi(x)-\frac{2+\log 2+o(1)}{\log\log x} +O(\xi^3(x))+O(\lambda(x)\xi^2(x))+O\left(\frac{\xi(x)}{\log\log x}\right).$$

Finding the maximum value of $f(p)$ boils down to finding the minimum value of $\Delta(x)$ and thus of $\xi^2(x)-\lambda(x)\xi(x)-\frac{2+\log 2+o(1)}{\log\log x}$. It is clear that this minimum is attained when $\xi(x)=\frac{\lambda(x)}{2}+o\left(\frac{1}{\log\log x}\right)$, in which case

$$\Delta(x)=2+\lambda(x)-\frac{\lambda^2(x)}{4}-\frac{2+\log 2+o(1)}{\log\log x}=2+\lambda(x)-\frac{2+\log 2+o(1)}{\log\log x},$$

thereby establishing (3.1) and (3.2).

Theorem 2. *Let $k\ge 2$ be a fixed integer and let*

$$f_k(p,x)=\#\{n\le x : P_k(n)=p\}.$$

Then, for large x, the maximum value of $f_k(p,x)$ is reached when $p=3$.

Proof. First consider the case $k=2$. We shall make use of the Prime Number Theorem in the form

$$\pi(x)=\frac{x}{\log x}+\frac{x}{\log^2 x}+O\left(\frac{x}{\log^3 x}\right).$$

In particular, one can easily show that

$$\pi(x/a)=\frac{1}{a}\frac{x}{\log x}+\frac{1+\log a}{a}\frac{x}{\log^2 x}+O\left(\frac{x}{\log^{5/2} x}\right), \tag{3.11}$$

uniformly for $1\le a\le \varphi(x)\stackrel{\text{def}}{=}\exp\{(\log x)^{1/4}\}$.

Let x be a large number, then, for $p=2$, we have

$$\begin{aligned} f_2(p,x) &= f_2(2,x)=\sum_{\substack{n\le x\\ P_2(n)=2}} 1=\sum_{\substack{2^m q\le x\\ m\ge 1}} 1=\sum_{m\ge 1}\pi\left(\frac{x}{2^m}\right)\\ &= \sum_{m\le M}\pi\left(\frac{x}{2^m}\right)+\sum_{m>M}\pi\left(\frac{x}{2^m}\right)=\Sigma_1+\Sigma_2, \end{aligned} \tag{3.12}$$

say, where $M=\left[\frac{(\log x)^{1/4}}{\log 2}\right]$ and where the second sum in (3.12) runs over primes q; here $[z]$ denotes the number of positive integers $\le z$. We now evaluate Σ_1 and Σ_2 separately.

First of all, we have trivially

$$\Sigma_2\le x\sum_{m>M}\frac{1}{2^m}=\frac{x}{2^M}\ll\frac{x}{\varphi(x)}\ll\frac{x}{\log^3 x}. \tag{3.13}$$

On the other hand, using (3.11) with $a=2^m$, for $m=1,2,\dots,M$, we obtain

$$\begin{aligned} \Sigma_1 &= \frac{x}{\log x}\sum_{m\le M}\frac{1}{2^m}+\frac{x}{\log^2 x}\sum_{m\le M}\frac{1+\log 2^m}{2^m}+O\left(\frac{x}{\log^{5/2}x}\right)\\ &= \frac{x}{\log x}\left(1-\sum_{m=M+1}^{\infty}\frac{1}{2^m}\right)\\ &\quad +\frac{x}{\log^2 x}\left(1+2\log 2-\sum_{m=M+1}^{\infty}\frac{1+\log 2^m}{2^m}\right)+O\left(\frac{x}{\log^{5/2}x}\right)\\ &= \frac{x}{\log x}+(1+2\log 2)\frac{x}{\log^2 x}+O\left(\frac{x}{\log^{5/2}x}\right), \end{aligned} \tag{3.14}$$

since

$$\sum_{m=M+1}^{\infty}\frac{1}{2^m}=\frac{1}{2^M}\ll\frac{1}{\varphi(x)}\ll\frac{1}{\log^3 x}$$

and similarly

$$\sum_{m=M+1}^{\infty}\frac{1+\log 2^m}{2^m}\ll\frac{1}{\varphi(x)}\ll\frac{1}{\log^3 x}.$$

Thus, using (3.13) and (3.14), (3.12) becomes

$$f_2(2,x)=\frac{x}{\log x}+(1+2\log 2)\frac{x}{\log^2 x}+O\left(\frac{x}{\log^{5/2}x}\right). \tag{3.15}$$

Now consider $p=3$. We then have

$$f_2(p,x)=f_2(3,x)=\sum_{\substack{2^r3^m q\le x\\ r\ge 0,\ m\ge 1}} 1=\sum_{r\ge 0,\ m\ge 1}\pi\left(\frac{x}{2^r3^m}\right)+O\left(\log^2 x\right),$$

where q runs over primes ≥ 3. Proceeding essentially in the same manner as with $p = 2$, we obtain

$$(3.16)\qquad f_2(3,x) = \frac{x}{\log x}\sum_{r=0}^{\infty}\sum_{m=1}^{\infty}\frac{1}{2^r 3^m} + \frac{x}{\log^2 x}\sum_{r=0}^{\infty}\sum_{m=1}^{\infty}\frac{1+\log(2^r 3^m)}{2^r 3^m} + O\left(\frac{x}{\log^{5/2} x}\right)$$
$$= \frac{x}{\log x} + (1 + \log 2 + \frac{3}{2}\log 3)\frac{x}{\log^2 x} + O\left(\frac{x}{\log^{5/2} x}\right).$$

Comparing (3.15) and (3.16), it follows that, if x is large enough, one has $f_2(3,x) > f_2(2,x)$ since $1 + \log 2 + \frac{3}{2}\log 3 > 1 + 2\log 2$, that is, 3 is more often found than 2 as the second largest prime divisor of an integer.

It remains to show that $f_2(p,x) < f_2(3,x)$ if $p \geq 5$. Hence, let p_ν denote the ν^{th} prime ($\nu \geq 3$). Proceeding as above, we have

$$f_2(p_\nu, x) = \sum_{\substack{2^{r_1}3^{r_2}\dots p_\nu^{r_\nu} q \leq x \\ r_1,r_2,\dots,r_{\nu-1}\geq 0,\ r_\nu \geq 1}} 1 = \sum_{\substack{r_1,r_2,\dots,r_{\nu-1}\geq 0 \\ r_\nu\geq 1}} \sum_{p_\nu \leq q \leq \frac{x}{2^{r_1}3^{r_2}\dots p_\nu^{r_\nu}}} 1$$
$$= \sum_{\substack{r_1,r_2,\dots,r_{\nu-1}\geq 0 \\ r_\nu\geq 1}} \pi\left(\frac{x}{2^{r_1}3^{r_2}\dots p_\nu^{r_\nu}}\right) + O\left(\log^\nu x\right),$$

where the first sum runs over primes $q \geq p_\nu$. Now observe that the series $\sum \frac{1}{2^{r_1}3^{r_2}\dots p_\nu^{r_\nu}}$ is convergent. Hence, using a weaker form of (3.11) and given any $\varepsilon > 0$, we may then write that, for $x \geq x_0(\varepsilon)$,

$$f_2(p_\nu, x) \leq (1+\varepsilon)\frac{x}{\log x}\sum_{\substack{r_1,r_2,\dots,r_{\nu-1}\geq 0 \\ r_\nu\geq 1}}\frac{1}{2^{r_1}3^{r_2}\dots p_\nu^{r_\nu}}$$
$$= (1+\varepsilon)\frac{x}{\log x}\sum_{r_1\geq 0}\frac{1}{2^{r_1}}\sum_{r_2\geq 0}\frac{1}{3^{r_2}}\cdots\sum_{r_{\nu-1}\geq 0}\frac{1}{p_{\nu-1}^{r_{\nu-1}}}\sum_{r_\nu\geq 1}\frac{1}{p_\nu^{r_\nu}}$$
$$= (1+\varepsilon)\frac{x}{\log x}2\cdot\frac{3}{2}\cdot\frac{5}{4}\cdots\frac{p_{\nu-1}}{p_{\nu-1}-1}\frac{1}{p_\nu - 1},$$

which can easily be shown to be smaller than $\frac{3}{4}\frac{x}{\log x}$, for all $\nu \geq 3$.

The case $k \geq 3$ can be handled in a similar manner.

Let Q be a set of primes satisfying (2.7) and denote by q_1 the smallest element of Q and by q_2 its second smallest element. Using essentially the same reasoning, one can prove the following result.

Theorem 3. *Let* $f_Q(p,x) \stackrel{\text{def}}{=} \#\{n \leq x : P(n,Q) = p\}$. *Then, for large* x, *the maximum value of* $f_Q(p,x)$ *is reached when* $p = q_1$, *unless* $q_1 = 2$ *and* $q_2 = 3$, *in which case the maximum is reached at* $p = 3$.

References

[1] K. Alladi and P. Erdős, *On an additive arithmetic function*, Pacific J. Math. **71** (1977), 275-294.

[2] K. Alladi and P. Erdős, *On the asymptotic behaviour of large prime factors of an integer*, Pacific J. Math. **82** (1979), 295-315.

[3] J.M. De Koninck, *Sur les plus grands facteurs premiers d'un entier*, Monatshefte Math. **116**, 13-37 (1993).

[4] J.M. De Koninck and J. Galambos, *The intermediate prime divisors of an integer*, Proc. Amer. Math. Soc. **101** (1987), 213-216.

[5] J.M. De Koninck and A. Ivić, *The distribution of the average prime divisor of an integer*, Archiv der Math. **43** (1984), 37-43.

[6] J.M. De Koninck and A. Ivić, *Arithmetic functions defined on sets of primes of positive density*, (to appear).

[7] P. Erdős and A. Ivić, *On sums involving reciprocals of certain arithmetical functions*, Publ. Inst. Math. Belgrade **32** (1982), 49-56.

[8] P. Erdős, A. Ivić and C. Pomerance, *On sums involving reciprocals of the largest prime factor of an integer*, Glasnik Matematički **21** (1986), 27-44.

[9] J. Galambos, *The sequences of prime divisors of an integer*, Acta Arith. **31** (1976), 213-218.

[10] D.A. Goldston and K.S. McCurley, *Sieving the positive integers by large primes*, J. Number Theory **28** (1988), 94-115.

[11] D.A. Goldston and K.S. McCurley, *Sieving the positive integers by small primes*, Trans. AMS **307** (1988), 51-62.

[12] J.L. Hafner and K.S. McCurley, *On the distribution of running times of certain integer factoring algorithms*, IBM Research Report, 1987.

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

[14] A. Ivić, *On the k-th prime factor of an integer*, Univ. u Novom Sadu Zb. Rad. Prirodno.-Mat. Fak. Ser. Mat. **20** (1990), 63-73.

[15] A. Ivić, *On large additive functions on primes of positive density*, (to appear).

[16] D.E. Knuth and L.T. Pardo, *Analysis of a simple factorization algorithm*, Theoretical Computer Science **3** (1976), 321-348.

[17] H. Riesel, *Prime Numbers and Computer Methods for Factorization*, Birkhäuser, 1985.

[18] G. Tenenbaum, *Introduction à la théorie analytique et probabiliste des nombres*, Institut Elie Cartan, 1990.

[19] M.C. Wunderlich and J.L. Selfridge, *A design for a number theory package with an optimized trial division routine*, Communications of the ACM **17** (1974), 272-276.

Jean-Marie De Koninck
Département de mathématiques et de statistique
Université Laval
Québec G1K 7P4
CANADA