THE MEDIAN LARGEST PRIME FACTOR

ERIC NASLUND

Abstract. Let \( M(x) \) denote the median largest prime factor of the integers in the interval \([1, x]\). We prove that

\[
M(x) = x^{1/\sqrt{e}} \exp(-\text{li}_f(x)/x) + O_\epsilon \left(x^{1/\sqrt{e}} e^{-c (\log x)^{3/5-\epsilon}}\right),
\]

where \( \text{li}_f(x) = \int_2^x \frac{(x/t)}{\log t} dt \). From this, we obtain the asymptotic

\[
M(x) = e^{\frac{\gamma - 1}{\sqrt{e}}} x^{1/\sqrt{e}} \left(1 + O \left(\frac{1}{\log x}\right)\right),
\]

where \( \gamma \) is the Euler-Mascheroni constant. This answers a question posed by Martin [3], and improves a result of Selfridge and Wunderlich [7].

1. Introduction

The median largest prime factor of the integers in the interval \([1, x]\), which we denote by \( M(x) \), has size

\[(1.1) \quad M(x) = x^{1/\sqrt{e} + o(1)}.\]

This result first appeared in 1974 in a paper by Selfridge and Wunderlich [7]. Martin asked [3] how the median prime factor compares to \( x^{1/\sqrt{e}} \), and whether we have

1. For sufficiently large \( x \), \( M(x) < x^{1/\sqrt{e}} \).
2. Each inequality \( M(x) < x^{1/\sqrt{e}} \) and \( M(y) > y^{1/\sqrt{e}} \) holds for arbitrarily large \( x, y \).
3. For sufficiently large \( x \), \( M(x) > x^{1/\sqrt{e}} \).

In this paper we prove that

\[(1.2) \quad M(x) = e^{\frac{\gamma - 1}{\sqrt{e}}} x^{1/\sqrt{e}} \left(1 + O \left(\frac{1}{\log x}\right)\right),\]

where \( \gamma \) is the Euler-Mascheroni constant. As \( \exp \left(\frac{\gamma - 1}{\sqrt{e}}\right) \approx 0.7738 \), it follows that option 1 holds.

Our main theorem is stronger than this with an error term like that of the prime number theorem, from which we may obtain an asymptotic expansion for \( M(x) \).

Theorem 1. For every \( \epsilon > 0 \),

\[(1.3) \quad M(x) = x^{1/\sqrt{e}} \exp(-\text{li}_f(x)/x) + O_\epsilon \left(x^{1/\sqrt{e}} e^{-c (\log x)^{3/5-\epsilon}}\right),\]

where \( \text{li}_f(x) = \int_2^x \frac{(x/t)}{\log t} dt \), and \( O_\epsilon \) means the constant is allowed to depend on \( \epsilon \).

The function \( \text{li}_f(x) \) admits an asymptotic expansion with coefficients expressible as sums of the Stieltjes constants.

Date: July 25th 2012.
Proposition 2. For any integer $k$, we have the asymptotic expansion
\begin{equation}
\text{li}_f(x) = c_0 \frac{x}{\log x} + c_1 \frac{x}{\log^2 x} + \cdots + c_{k-1} \frac{(k-1)!x}{\log^k x} + O\left(\frac{x}{\log^{k+1} x}\right),
\end{equation}
where $c_n = 1 - \sum_{k=0}^{n} \frac{1}{k!} \gamma_k$, and $\gamma_k$ denotes the $k^{th}$ Stieltjes.

Using the Taylor expansion for $e^z$ around $z = 0$ along with the above proposition, it follows that
\[
\exp\left(\frac{-\text{li}_f(x)}{x}\right) = 1 + \frac{1 - \gamma}{\log x} + O\left(\frac{1}{\log^2 x}\right).
\]
Since
\[
x^{\frac{1}{\sqrt{e}}} = e^{\frac{1}{\sqrt{e}} \frac{1 - \gamma}{x}}
\]
we are able to deduce equation (1.2) as a corollary of theorem 1. Applying the same approach with more terms, it follows that $M(x)$ has an asymptotic expansion of the form
\begin{equation}
M(x) = e^{\frac{1}{\sqrt{e}} x \frac{1}{\sqrt{e}} \left(1 + \frac{d_1}{\log x} + \cdots + \frac{d_n}{\log^n x} + O\left(\frac{1}{\log^{n+1} x}\right)\right)},
\end{equation}
where the $d_i$ are computable constants.

Our proof of theorem 1 is elementary, and uses an application of the hyperbola method. In section 3, we use theorem 1 to strengthen a result of Diaconis [2], and prove that
\begin{equation}
\sum_{n \leq x} \omega(n) = x \log x + B_1 x - \text{li}_f(x) + O_{\epsilon} \left(x e^{-(\log x)^{3/5 - \epsilon}}\right).
\end{equation}
From this, we can recover Diaconis' asymptotic expansion of $\sum_{n \leq x} \omega(n)$ by applying proposition 2. In section 4, we use the work of DeBruijn [1] and Saias [6] on integers without large prime factors to give an alternate derivation of theorem 1.

2. The Main Theorem

For each prime $p$ greater than $\sqrt{x}$, there is at most one integer $n \leq x$ such that $p|n$. By the definition of the median largest prime factor, exactly half of the integers in the interval $[1, x]$ will be divisible by a prime $p > M(x)$, and since $M(x) = x^{1/\sqrt{e} + o(1)} > \sqrt{x}$ there will be no double counting. It then follows that
\[
\frac{1}{2} x = \sum_{M(x) < p \leq x} \left\lfloor \frac{x}{p} \right\rfloor + O(1),
\]
where the $O(1)$ term arises since $x$ may not be an even integer. We may split up this sum by writing the floor function as $[x] = x - \{x\}$, where $\{x\}$ denotes the fractional part of $x$. Using Mertens formula
\begin{equation}
\sum_{p \leq x} \frac{1}{p} = \log \log x + B_1 + O_{\epsilon} \left(e^{-(\log x)^{3/5 - \epsilon}}\right),
\end{equation}
where $B_1 = \gamma - \sum_p \sum_{k \geq 2} \frac{1}{k p^k}$, we see that
\begin{equation}
\sum_{M(x) < p \leq x} \left\lfloor \frac{x}{p} \right\rfloor = x \left(\log \frac{\log x}{\log M(x)} - \frac{1}{2}\right) + O(1).
\end{equation}
To understand the left hand side of the equation, we need to find a precise asymptotic for the sum of the fractional parts \(\{x/p\}\). The following proposition strengthens a result of De La Vallée Poussin’s [5] where he gave the asymptotic \(\sum_{p \leq x} \{x/p\} \sim \frac{1 - \gamma}{\log x}\).

**Proposition 3.** We have that

\[
\sum_{p \leq x} \{\frac{x}{p}\} = li_f(x) + O_e \left( xe^{-c(\log x)^{3/5-\epsilon}} \right).
\]

**Proof.** Let \(1 < B \leq x\) be some integer, and fix \(\epsilon > 0\). Splitting into intervals and rearranging we have that

\[
\sum_{\frac{x}{B} < p \leq x} \left\{\frac{x}{p}\right\} = \sum_{n \leq B-1} n \left( \sum_{\frac{x}{n+1} < p \leq \frac{x}{n}} 1 \right) = \pi(x) + \pi \left( \frac{x}{2} \right) + \cdots + \pi \left( \frac{x}{B-1} \right) - (B-1)\pi \left( \frac{x}{B} \right)
\]

(2.3) \[
= \sum_{n \leq B-1} \left( \pi \left( \frac{x}{n} \right) - \pi \left( \frac{x}{B} \right) \right).
\]

By the prime number theorem this is

\[
= \sum_{n \leq B-1} \int_{\frac{x}{n+1}}^{\frac{x}{n}} \frac{1}{\log t} \, dt + O \left( \sum_{n \leq B-1} \left( \frac{x}{n} e^{-c\sqrt{\log \frac{x}{n}}} \right) \right)
\]

\[
= \int_{\frac{x}{B}}^{x} \frac{\{x/t\}}{\log t} \, dt + O \left( xe^{-c(\log x)^{3/5-\epsilon}} \log B \right).
\]

Using the fact that \([x] = x - \{x\}\), the main term is

\[
\int_{\frac{x}{B}}^{x} \frac{\{x/t\}}{\log t} \, dt = x \left( \log \log x - \log \log \left( \frac{x}{B} \right) \right) - \int_{\frac{x}{B}}^{x} \frac{\{x/t\}}{\log t} \, dt,
\]

and hence since \(\sum_{\frac{x}{B} \leq p \leq \frac{x}{B-1}} \frac{1}{p} = \log \log x - \log \log \left( \frac{x}{B} \right) + O \left( \left( \log x \right)^{3/5-\epsilon} \right)\) by [2.1], we have that

\[
\sum_{\frac{x}{B} < p \leq x} \left\{\frac{x}{p}\right\} = \int_{\frac{x}{B}}^{x} \frac{\{x/t\}}{\log t} \, dt + O \left( xe^{-c(\log x)^{3/5-\epsilon}} \log B \right).
\]

The proposition follows by choosing \(B = \sqrt{x}\) and noting that we can extend the sum and integral to start at 2 as \(\int_{2}^{\sqrt{x}} \frac{\{x/t\}}{\log t} \, dt = O \left( \frac{\sqrt{x}}{\log x} \right)\) and \(\sum_{p \leq \sqrt{x}} \left\{ \frac{x}{p} \right\} = O \left( \frac{\sqrt{x}}{\log x} \right)\). \(\square\)

Combining proposition 3 with equation 2.2, we have

\[
li_f(x) = -x \left( \log \frac{\log M(x)}{\log x} + \frac{1}{2} \right) + O_e \left( xe^{-c(\log x)^{3/5-\epsilon}} \right),
\]

since

\[
\sum_{M(x) < p \leq x} \left\{ \frac{x}{p} \right\} = \sum_{p \leq x} \left\{ \frac{x}{p} \right\} + O(M(x)) = li_f(x) + O_e \left( xe^{-c(\log x)^{3/5-\epsilon}} \right).
\]

Rearranging the equation, we find that

\[
\frac{\log M(x)}{\log x} = \exp \left( -\frac{1}{2} - \frac{\text{li}_f(x)}{x} + O_e \left( e^{-c(\log x)^{3/5-\epsilon}} \right) \right),
\]
and we are able to turn the error term into an additive factor since
\begin{equation}
\exp\left( O \left( e^{-c \left( \log x \right)^{3/5}} \right) \right) = 1 + O \left( e^{-c \left( \log x \right)^{3/5}} \right) .
\end{equation}

It follows that
\begin{equation}
M(x) = x^{1/\sqrt{x}} e^{-\operatorname{li}(x)/x} + O \left( \exp\left( -c \left( \log x \right)^{3/5} \right) \right) ,
\end{equation}
and by using \((2.5)\) again, we obtain equation \(1.3\) proving theorem \(1\).

2.1. The Function \(\operatorname{li}(x)\). To prove proposition \(1.4\), we begin by truncating the interval of integration to obtain
\begin{equation}
\operatorname{li}(x) = \int_{\sqrt{x}}^{x} \frac{x/t}{\log t} dt + O \left( \sqrt{x} \log x \right) .
\end{equation}

Substituting \(u = x/t\), we may write
\begin{equation}
\int_{\sqrt{x}}^{x} \frac{x/t}{\log t} dt = x \int_{1}^{\sqrt{x}} \frac{u}{u^2 \log \left( \frac{x}{u} \right)} du = x \int_{1}^{\sqrt{x}} \frac{u}{u^2} \left( 1 - \frac{\log u}{\log x} \right)^{-1} du .
\end{equation}

Expanding the geometric series
\begin{equation}
\left( 1 - \frac{\log u}{\log x} \right)^{-1} = 1 + \frac{\log u}{\log x} + \cdots + \left( \frac{\log u}{\log x} \right)^{k-1} + \left( \frac{\log u}{\log x} \right)^k \left( 1 - \frac{\log u}{\log x} \right)^{-1} ,
\end{equation}
we see that
\begin{equation}
\int_{\sqrt{x}}^{x} \frac{x/t}{\log t} dt = \frac{x}{\log x} \sum_{n=0}^{k-1} \int_{1}^{\sqrt{x}} \frac{u}{u^2} \left( \log u \right)^n du + \frac{x}{\log x} \int_{1}^{\sqrt{x}} \frac{u}{u^2} \left( \log u \right)^k \left( 1 - \frac{\log u}{\log x} \right)^{-1} du .
\end{equation}

The last term contributes an error of the form \(O_k \left( \frac{x}{\log^{k+1} x} \right)\), and since we may bound the integral
\begin{equation}
\int_{\sqrt{x}}^{x} u^2 \left( \log u \right)^n du \leq \int_{\sqrt{x}}^{x} \frac{\left( \log u \right)^n}{u^2} du = O \left( \frac{\left( \log x \right)^n}{\sqrt{x}} \right) ,
\end{equation}
it follows that by \(2.7\) and \(2.6\) we have
\begin{equation}
\operatorname{li}(x) = \frac{x}{\log x} \sum_{n=0}^{k-1} \int_{1}^{\sqrt{x}} \frac{u}{u^2} \left( \log u \right)^n du + O_k \left( \frac{x}{\log^{k+1} x} \right) .
\end{equation}

To evaluate the constants explicitly, we will make use of the Laurent expansion of \(\zeta(s)\) which is given by
\begin{equation}
\zeta(s) = \frac{1}{s-1} + \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \gamma_n (s-1)^n ,
\end{equation}
where the \(\gamma_n\) are the Stieltjes Constants, and \(\gamma_0\) is the Euler-Mascheroni constant. We will also make use of the identity
\begin{equation}
\zeta(s) = \frac{1}{s-1} + 1 - s \int_{1}^{\infty} \{x\} x^{-s-1} dx ,
\end{equation}
which holds for \(s \neq 1, \operatorname{Re}(s) > 0\) \([3]\). Letting
\begin{equation}
c_n = \frac{1}{n!} \int_{1}^{\infty} \frac{u}{u^2} \left( \log u \right)^n du ,
\end{equation}
we have the following lemma:

**Lemma 4.** For any integer \( n \geq 0 \),

\[
c_n = 1 - \sum_{k=0}^{n} \frac{1}{k!} \gamma_k.
\]

**Proof.** Consider the generating series

\[
\sum_{n=0}^{\infty} c_n z^n = \int_{1}^{\infty} \left\{ \frac{x}{u^2} \right\} e^{x \log u} du = \int_{1}^{\infty} \left\{ u \right\} u^{-2} du.
\]

By (2.10) this equals \( \frac{1 - \zeta(1 - 1/z)}{1 - z} \), and so from equation (2.9) we have

\[
\sum_{n=0}^{\infty} c_n z^n = \left( \sum_{m \geq 0} z^m \right) \left( 1 - \sum_{n \geq 0} \frac{\gamma_n}{n!} z^n \right),
\]

and the result follows upon comparing coefficients. \( \square \)

3. **The Mean of \( \omega(n) \) and \( \Omega(n) \)**

Using proposition 3, we are able to provide a short proof of the asymptotic expansion of \( \sum_{n \leq x} \omega(n) \) given in [2], where \( \omega(n) \) is the number of distinct prime divisors function. Since \( \omega(n) = \sum_{p|n} 1 \), rearranging the orders of summation implies that

\[
\sum_{n \leq x} \omega(n) = \sum_{n \leq x} \sum_{p|n} 1 = \sum_{p \leq x} x \left\{ \frac{x}{p} \right\} = x \sum_{p \leq x} \frac{1}{p} - \sum_{p \leq x} \left\{ \frac{x}{p} \right\}.
\]

By (2.1) and proposition 3, we deduce equation (1.6) which states that

\[
\sum_{n \leq x} \omega(n) = x \log \log x + B_1 x - \text{li}_f(x) + O \left( x e^{-c (\log x)^{3/5 - \epsilon}} \right).
\]

Combining the previous equation with proposition 2 yields Diaconis’ [2] expansion

\[
(3.1) \quad \sum_{n \leq x} \omega(n) = x \log \log x + B_1 x - \frac{c_0 x}{\log x} + \cdots + \frac{c_{k-1}(k-1)!x}{\log^k x} + O \left( \frac{x}{\log^{k+1} x} \right),
\]

with \( c_n \) given explicitly as \( c_n = 1 - \sum_{k=0}^{n} \frac{1}{k!} \gamma_k \). We may derive a similar expansion for \( \Omega(n) \), the number of distinct prime factors counted with multiplicity. Taking into account the higher prime powers, we see that

\[
\sum_{n \leq x} \Omega(n) - \omega(n) = x \sum_{p} \frac{1}{p(p-1)} + O \left( \sqrt{x \log x} \right),
\]

and so

\[
(3.2) \quad \sum_{n \leq x} \Omega(n) = x \log \log x + B_2 x - \text{li}_f(x) + O \left( x e^{-c (\log x)^{3/5 - \epsilon}} \right)
\]

where \( B_2 = B_1 + \sum_p \frac{1}{p(p-1)} \).
4. Integers Without Large Prime Factors

In this section, we deduce the main result in a different way using the work of DeBruijn [1] and Saias [6] on integers without large prime factors. Let \( \psi(x, y) \) denote the number of integers \( n \) with \( 1 \leq n \leq x \), all of whose prime factors are \( \leq y \). Then the median largest prime factor, \( M(x) \), of the integers in the interval \([1, x]\) satisfies

\[
\psi(x, M(x)) = \frac{1}{2} x + O(1).
\]

In [1], De Bruijn showed that

\[
\psi(x, y) = \Lambda(x, y) + O(\epsilon (xe^{\log x/3-5/3} - \epsilon)).
\]

where

\[
\Lambda(x, y) = x \int_0^x \rho \left( \frac{\log x - \log t}{\log y} \right) \frac{[t]}{t} dt,
\]

and \( \rho(u) \) denotes the Dickmann De Bruijn rho function. It follows that we are looking for \( y \) such that

\[
\Lambda(x, y) = \frac{1}{2} x + O(\epsilon (xe^{\log x/3-5/3} - \epsilon)).
\]

Examining \( \Lambda(x, y) \) more closely, integration by parts yields

\[
\int_0^x \rho \left( \frac{\log x - \log t}{\log y} \right) \frac{[t]}{t} dt = 1 + \frac{1}{\log y} \int_0^x \frac{[t]}{t^2} \rho' \left( \frac{\log x - \log t}{\log y} \right) dt.
\]

Substituting \( s = \frac{x}{t} \), and using the fact that \( \rho'(u) = 0 \) when \( 0 < u < 1 \), we have

\[
\Lambda(x, y) = x + \frac{1}{\log y} \int_y^x \frac{x}{s} \rho' \left( \frac{s}{\log y} \right) ds.
\]

In our case, since \( x > y > \sqrt{x} \), we are on the interval \( 1 < u < 2 \), and on this range \( \rho(u) = 1 - \log u \) so that \( \rho'(u) = -\frac{1}{u} \). Thus we have that for \( x \geq y > \sqrt{x} \),

\[
\Lambda(x, y) = 1 - \int_y^x \frac{x}{\log s} ds,
\]

and by splitting up the floor function and recalling the definition of \( \text{li}_f(x) \), it follows that

\[
\Lambda(x, y) = x - x \int_y^x \frac{1}{s \log s} ds + \int_y^x \frac{x}{s \log s} ds
\]

\[
= x \left( 1 - \log \left( \frac{\log x}{\log y} \right) \right) + \text{li}_f(x) + O(y).
\]

With [4.2] in hand, solving equation [4.1] for \( M(x) \) becomes identical to solving equation [2.4] and so this gives a way to deduce theorem [1] from De Bruijn’s work. To obtain proposition [2] we use the expansion for \( \Lambda(x, y) \) given in Saias’ paper [6]. Suppose that \( x^{\frac{1}{k}} = y, \ u \leq (\log y)^{3/5-\epsilon}, \) and that \( u \in \bigcup_{1 \leq k \leq n} (k + \epsilon, k + 1) \cup (n + 1, \infty) \), so that \( u \) is not too close to an integer. Then then we have the expansion

\[
\Lambda(x, y) = x \sum_{k=0}^n a_k \rho^{(k)}(u) \left( \frac{\log y}{\log y} \right)^k + O_{n, \epsilon} \left( x \frac{\rho^{(n+1)}(u)}{(\log y)^{n+1}} \right),
\]
where 

\[ a_0 = 1, \quad a_k = \frac{(-1)^k}{(k-1)!} \int_1^\infty \frac{\{t\}}{t^2} (\log t)^{k-1} dt. \]

In our case, \( 1 < u < 2 \), and on this range \( \rho(u) = 1 - \log u \) so that \( \rho'(u) = -\frac{1}{u} \) and \( \rho^{(k)}(u) = \frac{(-1)^k}{u^k} \). We note that, by \[4\] for \( k \geq 1 \)

\[ a_k = (-1)^k c_{k-1} = (-1)^k \left( 1 - \sum_{j=0}^{k-1} \frac{1}{j!} \gamma_j \right). \]

Since \( \frac{1}{(\log y)^k} = \frac{u^k}{(\log x)^k} \), equation \[4.3\] becomes

\[ (4.4) \quad \Lambda(x, y) = x\rho(u) + \sum_{k=1}^n \frac{(k-1)!c_{k-1}}{(\log x)^k} + O_n \left( \frac{x}{(\log x)^{n+1}} \right). \]

Combining equation \[4.4\] with the fact that \( \rho(u) = 1 - \log \left( \frac{\log x}{\log y} \right) \) when \( y > \sqrt{x} \), we are able to obtain the expansion in proposition \[2\] from equation \[4.2\]. We note that this implies that equations \[1.6\] and \[3.1\] on the mean of \( \omega(n) \) follow from the work of De Bruijn and Saias on integers without large prime factors.

Acknowledgments

I would like to thank Qiaochu Yuan for providing the proof of Lemma \[4\] on Math Overflow. I would also like to thank Andrew Granville for his insightful comments, and Kevin Ford for his helpful advice. I am indebted to Greg Martin for both his mathematical suggestions and his encouragement, as this project would not have started or progressed otherwise.

References

[1] N. G. de Bruijn. On the number of positive integers \( \leq x \) and free of prime factors \( > y \). \textit{Nederl. Acad. Wetensch. Proc. Ser. A.}, 54:50–60, 1951.

[2] P Diaconis. Asymptotic expansions for the mean and variance of the number of prime factors of a number \( n \), dept. Technical report, of Statistics Tech. Report 96, Stanford Univ, 1976.

[3] Greg Martin. Medium largest-prime-factor. MathOverflow, 2011. URL (accessed on 2011-03-04): http://mathoverflow.net/questions/57395.

[4] Hugh L. Montgomery and Robert C. Vaughan. \textit{Multiplicative number theory. I. Classical theory}, volume 97 of \textit{Cambridge Studies in Advanced Mathematics}. Cambridge University Press, Cambridge, 2007.

[5] C. De La Vallée Poussin. Sur les valeurs moyennes de certaines fonctions arithmétiques. \textit{Annales de la société scientifiques de Bruxelles}, 22:84–90, 1898.

[6] Éric Saias. Sur le nombre des entiers sans grand facteur premier. \textit{J. Number Theory}, 32(1):78–99, 1989.

[7] M. C. Wunderlich and J. L. Selfridge. A design for a number theory package with an optimized trial division routine. \textit{Comm. ACM}, 17(1):272–276, 1974.

E-mail address: naslund.eric@gmail.com

The University of British Columbia Department of Mathematics