COLLOQUIUM MATHEMATICUM

VOL. 120 2010 NO. 1

ON THE DISTRIBUTION OF THE EULER FUNCTION OF SHIFTED SMOOTH NUMBERS

BY

STEFANIE S. LOIPERDINGER and IGOR E. SHPARLINSKI (Sydney)

Abstract. We give asymptotic formulas for some average values of the Euler function on shifted smooth numbers. The result is based on various estimates on the distribution of smooth numbers in arithmetic progressions which are due to A. Granville and É. Fouvry & G. Tenenbaum.

1. Introduction. An integer $n \geq 1$ is called *y-smooth* if every prime factor p of n satisfies $p \leq y$. For a detailed introduction to smooth numbers, their properties and applications, see [1, 3, 4, 5, 7, 8, 9] and references therein.

We denote by $S(x, y)$ the set of numbers less than or equal to x that are y-smooth, that is,

$$S(x, y) = \{n : 1 \leq n \leq x \text{ and } n \text{ is } y\text{-smooth}\}.$$

Furthermore let $\Psi(x, y) = \#S(x, y)$ be the counting function for smooth numbers.

Also, as usual, we use $\varphi(k)$ to denote the Euler function of an integer $k \geq 1$.

In this paper, we obtain asymptotic formulas for some average values of the Euler function of shifted smooth numbers. Namely, for real $x \geq y \geq 2$, we define

$$T(x, y) = \sum_{\substack{a<n\leq x \\ n\in S(x,y)}} \frac{\varphi(n-a)}{n-a} \quad \text{and} \quad V(x, y) = \frac{1}{\Psi(x, y)} \sum_{\substack{a<n\leq x \\ n\in S(x,y)}} \varphi(n-a),$$

where $a \neq 0$ is a fixed integer (throughout the paper, the implied constant may depend on a).

2. Preparations. In what follows, we use $U = O(V)$, $U \ll V$, and $V \gg U$ as equivalents of the inequality $|U| \leq cV$ with some constant $c > 0$, which may depend only on n.

2010 *Mathematics Subject Classification*: 11N25, 11N37.
Key words and phrases: Euler function, smooth numbers.

DOI: 10.4064/cm120-1-10

We recall that the *Dickman–de Bruijn* function $\rho(u)$ is defined by

$$\rho(u) = 1, \qquad 0 \le u \le 1,$$

and

$$\rho(u) = 1 - \int_1^u \frac{\rho(v-1)}{v}\, dv, \qquad u > 1.$$

Then, by [9, Chapter III.5, Corollary 9.3], we have

LEMMA 1. *For any $\varepsilon > 0$, the estimate*

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

holds uniformly in the range

$$\exp((\log\log x)^{5/3+\varepsilon}) \le y \le x,$$

where

$$u = \frac{\log x}{\log y}.$$

The following asymptotic estimate on $\rho(u)$ follows immediately from a much more precise result of [9, Chapter III.5, Theorem 8].

LEMMA 2. *For any $u \to \infty$, we have*

$$\rho(u) = \exp(-(1+o(1))u\log u).$$

We note that the bound

$$\Psi(x,y) = xu^{-u+o(u)}, \tag{1}$$

due to Canfield, Erdős and Pomerance [1, Corollary to Theorem 3.1], holds in a much wider range than one can obtain from Lemmas 1 and 2; see also [5, 7, 9].

Furthermore, the following upper bound on the derivative of $\rho(u)$ is a very weak form of a much more precise result [9, Chapter III.5, Corollary 8.3].

LEMMA 3. *For any $u > 0$, we have*

$$\rho'(u) \ll \rho(u)\log(u+1).$$

For any integers a and d with $\gcd(a,d) = 1$, let

$$\Psi(x,y;a,d) = \#\{n \in S(x,y) : n \equiv a \pmod d\}$$

and let

$$\Psi_d(x,y) = \#\{n \in S(x,y) : \gcd(n,d) = 1\}.$$

In general, one expects that

$$\Psi(x, y; a, d) \sim \frac{\Psi_d(x, y)}{\varphi(d)}$$

for sufficiently large x.

Granville [3] has proved the following bounds on the average of smooth numbers lying in a fixed arithmetic progression.

LEMMA 4. *Let A be a fixed positive number. Then there exist positive constants γ and δ, depending only on A, such that for*

$$\Delta = \min\left\{\exp\left(\gamma \frac{\log y \log\log y}{\log\log\log y}\right), \frac{\sqrt{x}}{(\log x)^\delta}\right\}$$

uniformly over $y \ge 100$ we have

$$\sum_{d \le \Delta} \max_{z \le x} \max_{\gcd(a,d)=1} \left|\Psi(z, y; a, d) - \frac{\Psi_d(z, y)}{\varphi(d)}\right| = O\left(\frac{\Psi(x, y)}{(\log y)^A}\right),$$

where the implied constant depends only on A.

Finally, Fouvry and Tenenbaum [2] give the following asymptotic formula for the number of smooth numbers that are coprime to d.

LEMMA 5. *For any $\varepsilon > 0$ there exists $x_0(\varepsilon)$ such that for $x \ge x_0(\varepsilon)$, the estimate*

$$\Psi_d(x, y) = \frac{\varphi(d)}{d}\,\Psi(x, y)\left(1 + O\left(\frac{\log\log(dy)\log\log x}{\log y}\right)\right)$$

holds uniformly in the range

$$\exp((\log\log x)^{5/3+\varepsilon}) \le y \le x, \qquad \log\log(d+2) \le \left(\frac{\log y}{\log(u+1)}\right)^{1-\varepsilon},$$

where

$$u = \frac{\log x}{\log y}.$$

3. Asymptotic formulas. We are now ready to obtain our main results.

THEOREM 1. *There exists an absolute constant $C > 0$ such that for a sufficiently large x the bound*

$$T(x, y) = \Psi(x, y)\left(\frac{6}{\pi^2} + O\left(\frac{\log\log x \log\log y}{\log y}\right)\right)$$

holds uniformly in the range

$$x \ge y \ge \exp(C\sqrt{\log x \log\log\log x}).$$

Proof. Using the well known identity

$$\varphi(n) = n \sum_{d|n} \frac{\mu(d)}{d},$$

where $\mu(d)$ is the Möbius function (see [6, equation (16.3.1)]), and changing the order of summation, we can rewrite $T(x, y)$ in the following way:

$$T(x,y) = \sum_{\substack{a<n\le x\\ n\in S(x,y)}} \sum_{d|(n-a)} \frac{\mu(d)}{d} = \sum_{d\le x} \frac{\mu(d)}{d} \sum_{\substack{a<n\le x\\ n\in S(x,y)\\ n\equiv a \pmod d}} 1 = \sum_{d\le x} \frac{\mu(d)}{d} \Psi(x,y;a,d).$$

Let γ and δ be chosen as in Lemma 4, corresponding to $A = 1$. We now define

$$\Delta = \min\left\{\exp\left(\gamma \frac{\log y \log\log y}{\log\log\log y}\right), \frac{\sqrt{x/a}}{(\log x/a)^{\delta}}\right\},$$

and write

$$T(x,y) = \sum_{d\le x} \frac{\mu(d)}{d} \Psi(x,y;a,d) = \Sigma_1 + \Sigma_2, \tag{2}$$

where

$$\Sigma_1 = \sum_{d\le \Delta} \frac{\mu(d)}{d} \Psi(x,y;a,d); \qquad \Sigma_2 = \sum_{x\ge d>\Delta} \frac{\mu(d)}{d} \Psi(x,y;a,d).$$

For Σ_1 we have

$$\Sigma_1 = \sum_{d\le\Delta} \frac{\mu(d)\Psi_d(x,y)}{d\varphi(d)} + O(R), \tag{3}$$

where

$$R = \sum_{d\le\Delta} \frac{1}{d}\left|\Psi(x,y;a,d) - \frac{\Psi_d(x,y)}{\varphi(d)}\right|.$$

Now, for each divisor $f \mid a$, we collect together the terms with $\gcd(a, d) = f$, getting

$$R = \sum_{f|a} R_f, \tag{4}$$

where

$$\begin{aligned} R_f &= \sum_{\substack{d\le\Delta\\ \gcd(a,d)=f}} \frac{1}{d}\left|\Psi(x,y;a,d) - \frac{\Psi_d(x,y)}{\varphi(d)}\right| \\ &= \sum_{\substack{d\le\Delta\\ \gcd(a,d)=f}} \frac{1}{d}\left|\Psi(x/f,y;a/f,d/f) - \frac{\Psi_d(x,y)}{\varphi(d)}\right|, \end{aligned} \tag{5}$$

provided that $y > |a|$. We now note that Lemma 5 implies

$$\frac{\Psi_{d/f}(x/f,y)}{\varphi(d/f)} = \frac{1}{d/f}\,\Psi(x/f,y)\biggl(1+O\biggl(\frac{\log\log(dy)\log\log x}{\log y}\biggr)\biggr) \tag{6}$$
$$= \frac{f}{d}\,\Psi(x/f,y)\biggl(1+O\biggl(\frac{\log\log(dy)\log\log x}{\log y}\biggr)\biggr).$$

Furthermore, denoting

$$u_f = \frac{\log(x/f)}{\log y} = u + O\biggl(\frac{1}{\log y}\biggr),$$

where $u = \frac{\log x}{\log y}$, we see from Lemma 3 that

$$\rho(u_f) = \rho(u)\biggl(1+O\biggl(\frac{\log(u+1)}{\log y}\biggr)\biggr).$$

Thus, by Lemma 1 we have

$$\Psi(x/f,y) = \frac{1}{f}\,\Psi(x,y)\biggl(1+O\biggl(\frac{\log(u+1)}{\log y}\biggr)\biggr).$$

Therefore (6) can be rewritten as

$$\frac{\Psi_{d/f}(x/f,y)}{\varphi(d/f)} = \frac{1}{d}\,\Psi(x,y)\biggl(1+O\biggl(\frac{\log\log(dy)\log\log x}{\log y}+\frac{\log(u+1)}{\log y}\biggr)\biggr)$$
$$= \frac{1}{d}\,\Psi(x,y)\biggl(1+O\biggl(\frac{\log\log(dy)\log\log x}{\log y}\biggr)\biggr)$$

(since $u \ll \log x$). Applying Lemma 5 again, we obtain

$$\frac{\Psi_d(x,y)}{\varphi(d)} = \frac{\Psi_{d/f}(x/f,y)}{\varphi(d/f)}\biggl(1+O\biggl(\frac{\log\log(dy)\log\log x}{\log y}\biggr)\biggr).$$

Accordingly, since the series

$$\sum_{\substack{d=1\\ \gcd(a,d)=f}}^{\infty} \frac{1}{d\varphi(d/f)} < \infty$$

converges, we now derive from (5) that

$$R_f = \sum_{\substack{d\le\Delta\\ \gcd(a,d)=f}} \frac{1}{d}\biggl|\Psi(x/f,y;a/f,d/f) - \frac{\Psi_{d/f}(x/f,y)}{\varphi(d/f)}\biggr|$$
$$+O\biggl(\Psi(x,y)\,\frac{\log\log(\Delta y)\log\log x}{\log y}\biggr)$$
$$\ll \sum_{\substack{d\le\Delta\\ \gcd(a,d)=f}} \biggl|\Psi(x/f,y;a/f,d/f) - \frac{\Psi_{d/f}(x/f,y)}{\varphi(d/f)}\biggr|$$
$$+\Psi(x,y)\,\frac{\log\log(\Delta y)\log\log x}{\log y}.$$

Moreover, in the considered range of x and y, for sufficiently large x, we have

$$y \le \Delta^3,$$

hence

$$\log\log(\Delta y) \le \log\log(\Delta^4) \le \log\left(\frac{4\gamma \log y \log\log y}{\log\log\log y}\right) = O(\log\log y).$$

Since γ and δ in the definition of Δ are chosen to correspond to $A = 1$ in Lemma 4, we obtain

$$R_f \ll \Psi(x,y)\frac{\log\log x \log\log y}{\log y},$$

which after substitution into (4) yields

$$R \ll \Psi(x,y)\frac{\log\log x \log\log y}{\log y}. \tag{7}$$

We see that for $d \le \Delta$ the condition of Lemma 5:

$$\log\log(d+2) \le \left(\frac{\log y}{\log(u+1)}\right)^{1-\varepsilon}$$

is satisfied (provided x is large enough), so we derive

$$\begin{aligned} &\sum_{d\le\Delta}\frac{\mu(d)\Psi_d(x,y)}{d\varphi(d)} \\ &\qquad = \Psi(x,y)\sum_{d\le\Delta}\frac{\mu(d)}{d^2}\left(1+O\left(\frac{\log\log(dy)\log\log x}{\log y}\right)\right) \\ &\qquad = \Psi(x,y)\left(\sum_{d\le\Delta}\frac{\mu(d)}{d^2} + O\left(\frac{\log\log x}{\log y}\sum_{d\le\Delta}\frac{\log\log(dy)}{d^2}\right)\right). \end{aligned} \tag{8}$$

We also have

$$\begin{aligned} \sum_{d\le\Delta}\frac{\mu(d)}{d^2} &= \sum_{d=1}^{\infty}\frac{\mu(d)}{d^2} + O\left(\sum_{d\ge\Delta}\frac{1}{d^2}\right) = \frac{1}{\zeta(2)} + O\left(\frac{1}{\Delta}\right) \\ &= \frac{6}{\pi^2} + O\left(\frac{1}{\Delta}\right), \end{aligned} \tag{9}$$

where $\zeta(s)$ is the Riemann zeta-function (see [6, Theorem 287 and equation (17.2.2)]). To estimate the error term in (8) we use the trivial inequality

$$\sum_{d\le\Delta}\frac{\log\log(dy)}{d^2} \le \sum_{d\le\Delta}\frac{\log\log(\Delta y)}{d^2} \ll \log\log(\Delta y) \ll \log\log y. \tag{10}$$

Thus, substituting (9) and (10) in (8), we derive

$$\sum_{d\le\Delta}\frac{\mu(d)\Psi_d(x,y)}{d\varphi(d)}=\Psi(x,y)\biggl(\frac{6}{\pi^2}+O\biggl(\frac{\log\log x\log\log y}{\log y}\biggr)\biggr). \tag{11}$$

Combining (7) and (11), we deduce from (3) that

$$\Sigma_1=\Psi(x,y)\biggl(\frac{6}{\pi^2}+O\biggl(\frac{\log\log x\log\log y}{\log y}\biggr)\biggr). \tag{12}$$

For Σ_2 we have the trivial estimate

$$\begin{aligned}|\Sigma_2|&\le\sum_{x\ge d>\Delta}\frac{1}{d}\sum_{\substack{a<n\le x\\ n\in S(x,y)\\ n\equiv a\ (\mathrm{mod}\ d)}}1\le\sum_{x\ge d>\Delta}\frac{1}{d}\sum_{\substack{a<n\le x\\ n\equiv a\ (\mathrm{mod}\ d)}}1\\&\le\sum_{x\ge d>\Delta}\frac{1}{d}(\lfloor x/d\rfloor+1)\le 2x\sum_{x\ge d>\Delta}\frac{1}{d^2}=O\biggl(\frac{x}{\Delta}\biggr).\end{aligned} \tag{13}$$

Substituting (12) and (13) in (2), we obtain

$$T(x,y)=\Psi(x,y)\biggl(\frac{6}{\pi^2}+O\biggl(\frac{\log\log x\log\log y}{\log y}\biggr)\biggr)+O\biggl(\frac{x}{\Delta}\biggr). \tag{14}$$

We now see from Lemmas 1 and 2 that for a sufficiently large C, under the condition

$$x\ge y\ge\exp(C\sqrt{\log x\log\log\log x}),$$

the bound (1) holds, and furthermore we have

$$\begin{aligned}\Psi(x,y)\,\frac{\log\log x\log\log y}{\log y}&\ge\Psi(x,y)\,\frac{1}{\log y}\\&\ge x\exp\biggl(-2\,\frac{\log x}{\log y}\log\frac{\log x}{\log y}-\log\log y\biggr)\\&\ge x\exp\biggl(-2\,\frac{\log x}{\log y}\log\log x\biggr)\\&\ge\max\biggl\{x\exp\biggl(-\gamma\,\frac{\log y\log\log y}{\log\log\log y}\biggr),x^{1/2}(\log x)^{\delta}\biggr\}\\&=\frac{x}{\Delta}.\end{aligned}$$

Therefore the term $O(x/\Delta)$ can be removed from (14), which concludes the proof. ∎

THEOREM 2. *There exists an absolute constant $C>0$ such that for a sufficiently large x the bound*

$$V(x,y)=\frac{3x}{\pi^2}+O\biggl(\frac{x\log\log x\log\log y}{\log y}\biggr)$$

holds uniformly in the range

$$x \ge y \ge \exp(C\sqrt{\log x \log\log\log x}).$$

Proof. Using partial summation (see [9, Chapter I.0, Theorem 1]), we can rewrite $V(x,y)$ in the following way:

$$\begin{aligned} V(x,y) &= \frac{1}{\Psi(x,y)} \sum_{\substack{a<n\le x\\ n\in S(x,y)\\ n\equiv a \,(\mathrm{mod}\, d)}} \frac{\varphi(n-a)}{n-a}(n-a) \\ &= \frac{1}{\Psi(x,y)}\Big(T(x,y)(x-a) - \int_1^x T(t,y)\,dt\Big). \end{aligned}$$

For

$$t \le x \quad \text{and} \quad y \ge \exp(C\sqrt{\log x \log\log\log x})$$

Theorem 1 implies that

$$T(t,y) = \Psi(t,y)\bigg(\frac{6}{\pi^2} + O\bigg(\frac{\log\log t \log\log y}{\log y}\bigg)\bigg).$$

Therefore

$$V(x,y) = \frac{1}{\Psi(x,y)}\bigg(xT(x,y) - \bigg(\frac{6}{\pi^2} + O\bigg(\frac{\log\log x\log\log y}{\log y}\bigg)\bigg)\int_1^x \Psi(t,y)\,dt\bigg).$$

Since $\Psi(t,y) \le \Psi(x,y)$, this simplifies as

$$\begin{aligned} V(x,y) &= \frac{1}{\Psi(x,y)}\Big(xT(x,y) - \frac{6}{\pi^2}\int_1^x \Psi(t,y)\,dt\Big) \\ &\quad + O\bigg(\frac{x\log\log x\log\log y}{\log y}\bigg). \end{aligned} \tag{15}$$

Let, as before,

$$u = \frac{\log x}{\log y}.$$

By Lemma 1 we have

$$\begin{aligned} \Psi(t,y) &= t\rho\bigg(\frac{\log t}{\log y}\bigg)\bigg(1 + O\bigg(\frac{\log(u+1)}{\log y}\bigg)\bigg) \\ &= t\rho\bigg(\frac{\log t}{\log y}\bigg) + O\bigg(x\rho(u)\,\frac{\log(u+1)}{\log y}\bigg) \\ &= t\rho\bigg(\frac{\log t}{\log y}\bigg) + O\bigg(\Psi(x,y)\,\frac{\log(u+1)}{\log y}\bigg). \end{aligned}$$

Thus from (15) we derive

$$V(x,y) = \frac{1}{\Psi(x,y)}\Big(xT(x,y) - \frac{6}{\pi^2}\, I(x,y)\Big) + O\bigg(\frac{x\log\log x\log\log y}{\log y}\bigg), \tag{16}$$

where

$$I(x,y) = \int_1^x t\rho\left(\frac{\log t}{\log y}\right) dt.$$

Using integration by parts, we derive

$$\begin{aligned}
I(x,y) &= \frac{1}{2}\int_1^x \rho\left(\frac{\log t}{\log y}\right) dt^2 = \frac{1}{2}\,x^2\rho\left(\frac{\log x}{\log y}\right) + O(1) - \frac{1}{2}\int_1^x t^2\, d\rho\left(\frac{\log t}{\log y}\right) \\
&= \frac{1}{2}\,x^2\rho(u) + O(1) - \frac{1}{2\log y}\int_1^x t\rho'\left(\frac{\log t}{\log y}\right) dt.
\end{aligned}$$

By Lemma 3 we have

$$\int_1^x t\rho'\left(\frac{\log t}{\log y}\right) dt \ll \int_1^x t\rho\left(\frac{\log t}{\log y}\right) dt \log(u+1) \ll I(x,y)\log(u+1).$$

Therefore

$$I(x,y) = \frac{1}{2}\,x^2\rho(u) + O\left(1 + I(x,y)\,\frac{\log(u+1)}{\log y}\right),$$

which, together with Lemma 1, implies

$$\begin{aligned}
I(x,y) &= \frac{1}{2}\,x^2\rho(u)\left(1 + O\left(\frac{\log(u+1)}{\log y}\right)\right) \\
&= \frac{1}{2}\,x\Psi(x,y)\left(1 + O\left(\frac{\log(u+1)}{\log y}\right)\right).
\end{aligned}$$

Inserting this asymptotic formula in (16) and using Theorem 1, we conclude the proof. ■

4. Remarks. Certainly, improving the error term or obtaining similar bounds in a wider range are natural directions for further investigation.

Studying average values of other number-theoretic functions on shifted smooth numbers, such as

$$\frac{1}{\Psi(x,y)} \sum_{\substack{a<n\le x \\ n\in S(x,y)}} \tau(n-a) \quad \text{and} \quad \frac{1}{\Psi(x,y)} \sum_{\substack{a<n\le x \\ n\in S(x,y)}} \omega(n-a),$$

where $\tau(m)$ and $\omega(m)$ are the number of positive integer divisors and the number of prime divisors of $m \ge 1$, respectively, is of ultimate interest too. However, investigating these sums may require a very different approach.

Acknowledgments. Research of the first author was supported by an International Macquarie University Research Excellence Scholarship. Research of the second author was supported by ARC grant DP0556431.

REFERENCES

[1] E. R. Canfield, P. Erdős and C. Pomerance, *On a problem of Oppenheim concerning "Factorisatio Numerorum"*, J. Number Theory 17 (1983), 1–28.

[2] É. Fouvry et G. Tenenbaum, *Entiers sans grand facteur premier en progressions arithmétiques*, Proc. London Math. Soc. 63 (1991), 449–494.

[3] A. Granville, *Integers, without large prime factors, in arithmetic progressions. I*, Acta Math. 170 (1993), 255–273.

[4] —, *Integers, without large prime factors, in arithmetic progressions. II*, Philos. Trans. Roy. Soc. London Ser. A 345 (1993), 349–362.

[5] —, *Smooth numbers: computational number theory and beyond*, in: Algorithmic Number Theory: Lattices, Number Fields, Curves and Cryptography, J. P. Buhler and P. Stevenhagen (eds.), Cambridge Univ. Press, 2008, 267–323.

[6] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford Univ. Press, 1979.

[7] A. Hildebrand and G. Tenenbaum, *Integers without large prime factors*, J. Théor. Nombres Bordeaux 5 (1993), 411–484.

[8] K. Soundararajan, *The distribution of smooth numbers in arithmetic progressions*, in: Anatomy of Integers, CRM Proc. Lecture Notes 46, Amer. Math. Soc., Providence, RI, 2008, 115–128.

[9] G. Tenenbaum, *Introduction to Analytic and Probabilistic Number Theory*, Cambridge Univ. Press, 1995.

Stefanie S. Loiperdinger, Igor E. Shparlinski
Department of Computing
Macquarie University
Sydney, NSW, 2109, Australia
E-mail: stefanie@ics.mq.edu.au
igor@ics.mq.edu.au

Received 13 July 2009;
revised 14 October 2009 (5246)