JOURNAL
de Théorie des Nombres
de BORDEAUX
anciennement Séminaire de Théorie des Nombres de Bordeaux

Florian LUCA et Ravindranathan THANGADURAI

On an arithmetic function considered by Pillai

Tome 21, nº 3 (2009), p. 695-701.

<http://jtnb.cedram.org/item?id=JTNB_2009__21_3_695_0>

Journal de Théorie des Nombres
de Bordeaux **21** (2009), 695-701

On an arithmetic function considered by Pillai

par Florian LUCA et Ravindranathan THANGADURAI

Résumé. Soit n un nombre entier positif et $p(n)$ le plus grand nombre premier $p \leq n$. On considère la suite finie décroissante définie récursivement par $n_1 = n$, $n_{i+1} = n_i - p(n_i)$ et dont le dernier terme, n_r, est soit premier soit égal à 1. On note $R(n) = r$ la longueur de cette suite. Nous obtenons des majorations pour $R(n)$ ainsi qu'une estimation du nombre d'éléments de l'ensemble des $n \leq x$ en lesquels $R(n)$ prend une valeur donnée k.

Abstract. For every positive integer n let $p(n)$ be the largest prime number $p \leq n$. Given a positive integer $n = n_1$, we study the positive integer $r = R(n)$ such that if we define recursively $n_{i+1} = n_i - p(n_i)$ for $i \geq 1$, then n_r is a prime or 1. We obtain upper bounds for $R(n)$ as well as an estimate for the set of n whose $R(n)$ takes on a fixed value k.

1. Introduction

Let $n > 1$ be an integer. Let $p(n)$ be the largest prime factor of n. Let $n_2 = n_1 - p(n_1)$. If $n_2 > 1$, let $n_3 = n_2 - p(n_2)$, and, recursively, if $n_k > 1$, we put $n_{k+1} = n_k - p(n_k)$. Note that if n_k is prime, then $n_{k+1} = 0$. We put $R(n)$ for the positive integer k such that n_k is prime or 1. Hence, we obtain a representation of n of the form

$$n = p_1 + p_2 + \cdots + p_r, \tag{1.1}$$

with $r = R(n)$, where $p_1 > p_2 > \cdots > p_r$ are primes except for the last one which might be 1.

The above representation of n was first considered by Pillai in [6] who obtained a number of interesting results concerning the function $R(n)$. Here, we extend some of Pillai's results on this function.

Since by Bertrand's postulate the interval $[x, 2x)$ contains a prime number for all $x \geq 1$, it follows that if $n_k > 1$, then $n_{k+1} \leq n_k/2$. This immediately implies that $R(n) = O(\log n)$. Pillai proved that the better estimate $R(n) = o(\log n)$ holds as $n \to \infty$. He also showed, under the

Manuscrit reçu le 26 novembre 2007.

Riemann Hypothesis, that the inequality $R(n) < 2\log\log n$ holds whenever $n > n_0$. Here, we remove the conditional assumption on the Riemann Hypothesis from Pillai's result and prove the following theorem.

Theorem 1.1. *The estimate*

$$R(n) \ll \log\log n$$

holds for all positive integers $n \geq 3$.

Pillai also showed that

$$\limsup_{n\to\infty} R(n) = \infty. \tag{1.2}$$

Our next result is slightly stronger than estimate (1.2) above. In what follows, we put $\log_k x$ for the function defined inductively as $\log_1 x = \log x$ and $\log_k x = \max\{1, \log(\log_{k-1} x)\}$ for $k > 1$. When $k = 1$, we omit the subscript. Note that if x is large, then $\log_k x$ coincides with the kth fold composition of the natural logarithm function evaluated in x.

Theorem 1.2. *Let* $k \geq 1$ *be any fixed integer. Then the estimate*

$$\#\{n \leq x \ : \ R(n) = k\} \ \asymp_k \ \frac{x}{\log_k x}$$

holds.

Theorem 1.2 shows that for any fixed k, the asymptotic density of the set of n with $R(n) \leq k$ is zero. This shows not only that estimate (1.2) holds, but that $R(n) \to \infty$ holds on a set of n of asymptotic density 1.

Pillai also conjectured that perhaps the inequality $R(n) \gg \log\log n$ holds for infinitely many n. We believe this conjecture to be false. Indeed, a widely believed conjecture of Cramér [2] from 1936, asserts that if $x > x_0$, then the interval $[x, x + (\log x)^2]$ contains a prime number. If true, this implies that if $n_k > x_0$, then $n_{k+1} < (\log n_k)^2$. Let $f(n)$ be the function which associates to each integer $n > x_0$ the minimal number of iterations of the function $x \mapsto (\log x)^2$ required to take n just below x_0. Then Cramér's conjecture implies that $R(n) \leq f(n) + O(1)$, where the constant implied in $O(1)$ can be taken to be $\max\{R(n) : n \leq x_0\}$. Let us take a look at these iterations. Assume that n is large. We then have $n_1 = n$, $n_2 \leq (\log n)^2$, $n_3 \leq (\log n_2)^2 \leq (2\log(2\log n))^2 < 8(\log\log n)^2$. Inductively, one shows that if k is fixed and n is sufficiently large with respect to k, then the inequality $n_k < 8(\log_k n)^2$ holds. Since k is arbitrary, we conclude that $f(n) = o(\log_k n)$ holds with any fixed $k \geq 1$ as $n \to \infty$, so, in particular, the inequality $f(n) \gg \log\log n$ cannot hold for infinitely many positive integers n. Let us observe that the weaker assumption that the interval $[x, x + \exp((\log x)^{1/2})]$ contains a prime for all $x > x_0$ will easily lead to the conclusion that $R(n) = O(\log_3 n)$. Indeed, in this case we have $\log n_{k+1} \le (\log n_k)^{1/2}$, whenever $n_k > x_0$. In particular, $\log n_{k+1} \le (\log n)^{1/2^k}$, whenever $n_{k+1} > x_0$. This implies easily that for some k of size at most $(\log\log\log n)/\log 2 + O(1)$ we have $n_{k+1} < x_0$, so that $R(n) = O(\log_3 n)$.

Pillai also looked at the sequence of local maxima (in modern terms also called *champions*) for the function $R(n)$. Recall that n is called a *champion* if $R(m) < R(n)$ holds for all $m < n$. Let $\{t_k\}_{k\ge1}$ be the sequence of champions. Pillai showed that $t_1 = 1$ and that the recurrence $t_{k+1} = p(t_{k+1}) + t_k$ holds for all $k \ge 1$. Furthermore, t_k and t_{k+1} have different parities for all $k \ge 1$. He also showed that $\{t_k\}_{k\ge1}$ grows very fast, namely that for each positive constant A one has $t_{k+1} \gg_A t_k(\log t_k)^A$. He also calculated the first 4 values of the sequence $\{t_k\}_{k\ge1}$ obtaining

$$t_1 = 1, \quad t_2 = 4 = 3 + 1, \quad t_3 = 27 = 23 + 4, \quad t_4 = 1354 = 1327 + 27.$$

He mentioned (seventy years ago!) that it is perhaps possible to compute t_5 but not t_6. Consulting Thomas Nicely's [5] tables of prime gaps, we get

$$t_5 = 401429925999155061 = 401429925999153707 + 1354$$

and Cramér's conjecture implies that $t_6 > \exp(4 \cdot 10^8)$, so indeed it is perhaps not possible to compute t_6.

2. Proof of Theorem 1.1

For the proof of the fact that $R(n) < 2\log\log n$ for $n > n_0$ under the Riemann Hypothesis, Pillai used the known consequence of the Riemann Hypothesis that for each $\delta > 0$, there is some x_δ such that when $x > x_\delta$, the interval $[x, x + x^{1/2+\delta}]$ contains a prime number.

In the same year as Pillai's paper [6] appeared, Hoheisel proved his famous theorem about Prime Number Gaps.

Theorem 2.1 ([4])**.** *There exist absolute constants $\theta \in (0,1)$ and N_0 such that for every integer $n \ge N_0$, the interval $[n - n^\theta, n]$ contains a prime number.*

The best known $\theta = 0.525$ is due to Baker, Harman and Pinz [1]. The proof of Theorem 1.1 follows easily from Pillai's arguments by replacing the prime number gaps guaranteed by the Riemann Hypothesis with Hoheisel's result.[1]

---

[1] It seems likely that Pillai was not aware of Hoheisel's paper [4].

Let $n_1 \geq N_0$. By Theorem 2.1, $p(n_1) > n - n^\theta$. Thus, the chain of inequalities

$$\begin{aligned} n_2 &= n_1 - p(n_1) < n_1 - n_1 + n_1^\theta = n_1^\theta; \\ n_3 &= n_2 - p(n_2) < n_2^\theta < n_1^{\theta^2}; \\ n_4 &< n_1^{\theta^3}; \\ &\cdots\cdots\cdots \\ n_{\ell+1} &< n_1^{\theta^\ell} \end{aligned}$$

holds as long as $n_\ell \geq N_0$. We now let ℓ be that integer such that $n_{\ell+2} < N_0 \leq n_{\ell+1}$. We then have

$$n_1^{\theta^\ell} \geq N_0,$$

therefore

$$\theta^\ell \log n_1 \geq \log N_0,$$

which implies that

$$\ell \log \theta + \log\log n_1 \geq \log\log N_0.$$

Hence,

$$\log\log n_1 \geq \ell \log(1/\theta),$$

which in light of the fact that $\theta \in (0,1)$ gives

$$\ell \leq \frac{\log\log n_1}{\log(1/\theta)}.$$

Put $b = \max_{1 \leq m \leq N_0}\{R(m)\}$. Trivially, $b \leq \pi(N_0)$. Thus,

$$R(n_1) \leq \ell + 1 + b < \frac{\log\log n_1}{\log(1/\theta)} + 1 + b \ll \log\log n_1,$$

which is the desired inequality.

3. Proof of Theorem 1.2

For every prime number p we put p' for the next prime following p. The following result is certainly well-known but we shall supply a short proof of it.

Lemma 3.1. *For* $2 \leq y \leq \log x$, *put*

$$\mathcal{P}(x,y) = \left\{p \leq x : p' - p \notin [y^{-1}(\log x), y(\log x)]\right\}.$$

Then,

$$\#\mathcal{P}(x,y) \ll \frac{\pi(x)}{y}. \tag{3.1}$$

Proof. We first look at the primes $p \le x$ which are in $\mathcal{P}(x,y)$ and $p' - p > y \log x$. The interval $[1, x]$ is contained in the union of the subintervals $[(i-1)y\log x, iy(\log x)]$ for $i = 1, 2, \ldots, \lfloor x/(y \log x)\rfloor + 1$. Since $p' - p > y(\log x)$, each one of the above intervals can contain at most one such prime p. Thus, the number of such primes p does not exceed

$$\begin{aligned} \#\{p \le x \ : \ p' - p > y(\log x)\} &\le \lfloor x/(y \log x\rfloor + 1 \le 2x/(y\log x) \\ &\ll \pi(x)/y. \end{aligned} \tag{3.2}$$

We next look at the primes $p \le x$ which are in $\mathcal{P}(x,y)$ and $p' - p = h < z = y^{-1}(\log x)$. We fix h and look at the set of primes $p \le x$ such that $p + h$ is also prime. We write $\mathcal{A}_h(x)$ for this set. By Brun's sieve (see, for example, [3, Theorem 5.7]), we have

$$\#\mathcal{A}_h(x) \ll \frac{x}{(\log x)^2}\frac{h}{\phi(h)}.$$

Summing up over all the acceptable values of $h \le z$, we get that

$$\begin{aligned} \#\{p \le x \ : \ p' - p < z\} &\le \sum_{1\le h\le z} \#\mathcal{A}_h \le \frac{x}{(\log x)^2} \sum_{1 \le h \le z} \frac{h}{\phi(h)} \\ &\ll \frac{xz}{(\log x)^2} \ll \frac{\pi(x)}{y}. \end{aligned} \tag{3.3}$$

In the above estimates, we used the known fact that the estimate

$$\sum_{1\le h\le t} \frac{h}{\phi(h)} \ll t$$

holds for all $t \ge 1$ (see, for example, [7]). The desired conclusion follows now immediately from estimates (3.2) and (3.3). $\square$

Proof of Theorem 1.2. We put $\mathcal{R}_k = \{n : R(n) = k\}$ and $\mathcal{R}_k(x) = \mathcal{R}_k \cap [1, x]$. We prove the theorem by induction on k having as a base the case $k = 1$ for which the assertion is immediate by the Prime Number Theorem.

Assume that $k \ge 2$. We first deal with the upper bound on $\#\mathcal{R}_k(x)$. We have, by the induction hypothesis,

$$\begin{aligned} \#\mathcal{R}_k(x) &= \#\{n = p + m \le x \ : \ R(m) = k - 1, \ p \le n < p'\} \\ &= \sum_{p\le x} \#\{m \le p' - p \ : \ R(m) = k-1\} \\ &\le \sum_{p \le x} \#\mathcal{R}_{k-1}(p'-p) \ll_k \sum_{p\le x} \frac{(p'-p)}{\log_{k-1}(p'-p)}. \end{aligned} \tag{3.4}$$

We split the last sum above at $z = (\log x)^{1/3}$. If $p' - p > z$, then $\log_{k-1}(p' - p) \gg_k \log_k x$, therefore

$$\sum_{\substack{p\le x\\ p'-p>z}} \frac{(p'-p)}{\log_{k-1}(p'-p)} \ll_k \frac{1}{\log_k x}\sum_{p\le x}(p'-p) \ll \frac{x}{\log_k x}, \tag{3.5}$$

where for the last inequality above we used the fact that the intervals $[p, p')$ for $p \le x$ are disjoint and their union is contained in $[1, 2x]$ by the Bertrand postulate. For the range $p' - p \le z$, we proceed as in the proof of Lemma 3.1 by first fixing $h \le z$ and looking at the primes $p \in \mathcal{A}_h(x)$. The proof of Lemma 3.1 shows that

$$\begin{aligned}\sum_{p\in\mathcal{A}_h(x)} \frac{(p'-p)}{\log_{k-1}(p'-p)} &\ll \sum_{p\in\mathcal{A}_h(x)} h \le h\#\mathcal{A}_h \ll \frac{x}{\log x}\frac{h^2}{\phi(h)}\\ &\ll \frac{xz}{\log x}\frac{h}{\phi(h)},\end{aligned}$$

therefore

$$\sum_{\substack{p\le x\\ p'-p\le z}} \frac{(p'-p)}{\log_{k-1}(p'-p)} \ll \frac{xz}{\log x}\sum_{h\le z}\frac{h}{\phi(h)} \ll \frac{xz^2}{\log x} = \frac{x}{z} \ll \frac{x}{\log_k x}. \tag{3.6}$$

Estimates (3.4), (3.5) and (3.6) imply the desired upper bound on $\#\mathcal{R}_k(x)$.

We now turn our attention on the lower bound for $\#\mathcal{R}_k(x)$. We proceed again by induction on $k \ge 1$. Let $c_1 > 0$ be the constant implied in inequality (3.1) and let $y = 2c_1$. Then $\#\mathcal{P}(x, y) \le \pi(x)/2$. Let $p \le x$ be a prime not in $\#\mathcal{P}(x, y)$ and $m \in \mathcal{R}_{k-1}((\log x)/y)$. Put $n = m+p$. Then $n = m+p < (\log x)/y+p < p'$, therefore $p = p(n)$. Thus, $R(n) = 1+R(m) = k$. The number of pairs (p, m) with the above properties is

$$\begin{aligned}&\ge (\pi(x) - \#\mathcal{P}(x, y))\#\mathcal{R}_{k-1}((\log x)/y) \gg_k \frac{\pi(x)\log x}{\log_{k-1}((\log x)/y)}\\ &\gg_k \frac{x}{\log_k x}.\end{aligned}$$

Each such pair (p, m) leads to a value of $n \le x + (\log x)/y \le 2x$. Furthermore, distinct pairs (p, m) lead to distinct values of n, for if $p+m = p_1+m_1$ for some $(p, m) \ne (p_1, m_1)$ then, assuming say that $p_1 > p$, we get

$$p' - p \le p_1 - p = m - m_1 < m < (\log x)/y,$$

which is impossible. Hence, $p_1 = p$ and since $p + m = p_1 + m_1$, we also get $m = m_1$, which is impossible since the pairs (p, m) and (p_1, m_1) were distinct. Thus, we showed that $\#\mathcal{R}_k(2x) \gg_k x/\log_k x$, which implies the desired lower bound.

Acknowledgements. We thank the referee for suggestions which improved the quality of the paper. Work on this paper started during a pleasant visit of F. L. to the Harish-Chandra Research Institute in August, 2007. This author thanks the people of this institute for their hospitality as well as the TWAS for financial support. Research of F. L. was also supported in part by grants SEP-CONACyT 79685 and PAPIIT 100508.

References

[1] R. C. Baker, G. Harman and J. Pintz, *The difference between consecutive primes - II.* Proc. London Math. Soc., (3) **83** (2001), 532–562.

[2] H. Cramér, *On the order of magnitude of the differences between consecutive prime numbers.* Acta. Arith., **2** (1936), 396–403.

[3] H. Halberstam and H. E. Rickert, *Sieve methods.* Academic Press, London, UK, 1974.

[4] G. Hoheisel, *Primzahlprobleme in der Analysis.* Sitzunsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin, **33** (1930), 3–11.

[5] T. R. Nicely, *Some Results of Computational Research in Prime Numbers.* http://www.trnicely.net/

[6] S. S. Pillai, *An arithmetical function concerning primes.* Annamalai University J. (1930), 159–167.

[7] R. Sitaramachandra Rao, *On an error term of Landau - II* in "Number theory (Winnipeg, Man., 1983)", Rocky Mountain J. Math. **15** (1985), 579–588.

Florian Luca
Mathematical Institute
UNAM, Ap. Postal
61-3 (Xangari), CP 58089
Morelia, Michoacán, Mexico
E-mail: fluca@matmor.unam.mx

Ravindranathan Thangadurai
Harish-Chandra Research Institute
Chhatnag Road, Jhunsi
Allahabad 211 019, India
E-mail: thanga@mri.ernet.in