Acta Math. Hungar., **128 (4)** (2010), 307–314.
DOI: 10.1007/s10474-010-9191-x
First published online March 18, 2010

PRIME-REPRESENTING FUNCTIONS

K. MATOMÄKI

Department of Mathematics, University of Turku, 20014 Turku, Finland
e-mail: ksmato@utu.fi

(Received August 24, 2009; revised November 13, 2009; accepted November 16, 2009)

Abstract. We construct prime-representing functions. In particular we show that there exist real numbers $\alpha > 1$ such that $\lfloor \alpha^{2^n} \rfloor$ is prime for all $n \in \mathbb{N}$. Indeed the set consisting of such numbers α has the cardinality of the continuum.

1. Introduction

A well-known question is whether there exist simple functions whose all values are distinct primes. Having such an explicit and easily calculable function would give us an infinite reserve of prime numbers. However, given the irregularity of the distribution of primes, it is hard to believe that such a function exists.

Prime-representing functions, that is functions whose all values are primes, have got some attention in the past, and there are some neat but non-practical examples. They typically include an unknown parameter α that depends on the prime sequence which the function represents. One cannot determine which values of α lead to prime-representing functions but it is possible to show that there exist such numbers.

Mills [5] showed in 1947 that there exists $\alpha > 1$ such that

$$\lfloor \alpha^{3^n} \rfloor \tag{1}$$

is prime for all $n \in \mathbb{N}$. Later Niven [6] showed that 3 in the exponent could be replaced by any real number

$$c > \frac{8}{3} = \frac{1}{1 - 5/8}.$$

Here 5/8 comes from Ingham's [3] result that, for some $C > 0$, the interval $\left[x, x + Cx^{5/8}\right]$ contains primes for every sufficiently large x. Ingham's result

Key words and phrases: prime representing function, distribution of primes.
2000 Mathematics Subject Classification: 11A41, 11N05.

has been improved several times since, the best one being the following result by Baker, Harman and Pintz [2]. There and later we write $\pi(x)$ for the number of primes below x.

LEMMA 1. *There exists a positive constant d_0 such that*

$$\pi\left(x + x^{21/40}\right) - \pi(x) \geqq d_0 \frac{x^{21/40}}{\log x}$$

for every sufficiently large x.

Niven's argument then gives that (1) still holds if 3 is replaced by any exponent

$$c \geqq \frac{1}{1 - 21/40} = \frac{40}{19} \approx 2.1053.$$

This was quite recently noticed by Alkauskas and Dubickas in [1], where they, among other theorems, showed essentially the following [1, Theorem 1].

THEOREM 2. *Let $c_i \geqq 2.1053$ for every $i \in \mathbb{N}$ and let $C_n = c_1 \cdots c_n$. Then there exists $\alpha > 1$ such that the sequence $\lfloor \alpha^{C_n} \rfloor$ contains only prime numbers. If, in addition, $\limsup_{n\to\infty} c_n = \infty$, then α can be chosen to be transcendental.*

We notice two refinements to Niven's result. Firstly this shows that, under a certain condition, there is a transcendental α which leads to a prime-representing sequence. Secondly this shows that the exponent function c^n can be replaced by a more general product $c_1 \cdots c_n$.

These two observation were already implicitly present fifty years earlier in Wright's paper [10] that developed the theory of representing functions. Wright showed that the set of possible numbers α has the cardinality of the continuum, is nowhere dense and has measure zero. The cardinality claim naturally already implies that there are transcendental choices for α even without assuming the condition on lim sup.

In this paper we follow the lines of Wright and prove the following theorem which extends the admissible range for c_i to $c_i \geqq 2$.

THEOREM 3. *Let $c_i \geqq 2$ for every $i \in \mathbb{N}$ and let $C_n = c_1 \cdots c_n$. Then there exists $\alpha > 2$ such that the sequence $\lfloor \alpha^{C_n} \rfloor$ contains only prime numbers. The set of such numbers α has the cardinality of the continuum, is nowhere dense and has measure zero.*

Taking $c_i = 2$ for every $i \in \mathbb{N}$, this implies the following corollary.

COROLLARY 4. *There exists $\alpha > 2$ such that the sequence $\lfloor \alpha^{2^n} \rfloor$ contains only prime numbers. The set of such numbers α has the cardinality of the continuum, is nowhere dense and has measure zero.*

The main new ingredient in this paper is the author's result in [4] on sums of differences between consecutive primes. We will need to redo some of Wright's work to be able to use that result.

2. ϕ-sequences

Let $\lambda_n(x) = x^{c_n}$. Let further $\phi_0(x) = x$ and $\phi_n(x)$ be the composed function

$$\phi_n(x) = \lambda_n \circ \cdots \circ \lambda_1(x) = x^{C_n}$$

for $n \in \mathbb{N}$.

We say that a sequence (a_n) of positive integers is a *ϕ-sequence* if, for some fixed $\alpha > 1$, $a_n = \lfloor \phi_n(\alpha) \rfloor$ for every $n \in \mathbb{N}$.

In [10] these notions are defined for more general functions $\lambda_n(x)$. Our choice of functions $\lambda_n(x)$ satisfies conditions there, and so we can apply results from [10] to this special case. The following lemma (which is [10, Theorem 2]) gives a sufficient condition for a sequence to be a ϕ-sequence.

LEMMA 5. *Assume that $a_0 > 2$,*

$$\lambda_{n+1}(a_n) \leqq a_{n+1} \leqq \lambda_{n+1}(a_n + 1) - 1$$

for all $n \in \mathbb{N}$ and

$$a_{n+1} < \lambda_{n+1}(a_n + 1) - 1$$

for infinitely many $n \in \mathbb{N}$. Then the sequence (a_n) is a ϕ-sequence.

What we need to do is to show that there is a prime sequence a_n which satisfies the conditions of this lemma.

More generally, we let $\mathcal{B}$ to be an infinite set of positive integers and $c \geqq 2$. We write $E_c(\phi, \mathcal{B})$ for the set of all $\alpha \geqq c$ such that $\lfloor \phi_n(\alpha) \rfloor \in \mathcal{B}$ for all $n \in \mathbb{N}$. We combine a series of results from Wright [10] into one lemma. The inequality

$$\lambda_n(m) < k < \lambda_n(m + 1) - 1 \tag{2}$$

will occur repeatedly.

LEMMA 6. *Let $c \geqq 2$.*

(i) *If, for every $n \geqq 1$ and every $m \geqq \phi_{n-1}(c)$, there exists $k \in \mathcal{B}$ such that* (2) *holds, then $E_c(\phi, \mathcal{B}) \neq \emptyset$.*

(ii) *Assume that the condition in* (i) *is satisfied. If there are infinitely many integers $n \geqq 1$ such that, for every integer $m \geqq \phi_{n-1}(c)$, there are at least two distinct $k, k' \in \mathcal{B}$ for which* (2) *holds, then $E_c(\phi, \mathcal{B})$ has the cardinality of the continuum.*

(iii) *If there are infinitely many integers n such that, for every integer $m \geqq \phi_{n-1}(c)$, there is at least one integer $k \notin \mathcal{B}$ for which* (2) *holds, then the set $E_c(\phi, \mathcal{B})$ is nowhere dense.*

(iv) *Assume that $\phi_n(x)$ is convex for every $n \in \mathbb{N}$ and that there exists an integer r such that out of every r consecutive positive integers at least one is not in $\mathcal{B}$. Then $E_c(\phi, \mathcal{B})$ is of zero measure.*

Proof. These are [10, Theorems 4–7]. The proof of (i) follows from Lemma 5 by taking any $a_0 \geqq c$ and then choosing numbers $a_n \in \mathcal{B}$ for $n \geqq 1$ recursively so that the conditions of Lemma 5 are satisfied. This is possible by using (2) at each stage with $m = a_{n-1}$.

Claim (ii) follows in the same way noticing that, for infinitely many n, we have at least two choices for a_n, so that there are $2^{\aleph_0}$ possible sequences (a_n). Each of these must correspond to different α, so the set $E_c(\phi, \mathcal{B})$ has the cardinality of the continuum.

For proofs of (iii) and (iv), see [10, Theorems 6 and 7]. $\square$

Using Lemma 1, one can now conclude the result of Alkauskas and Dubickas (Theorem 2) from Lemma 6 with $\mathcal{B}$ the set of prime numbers $\mathbb{P}$ and c a sufficiently large positive constant.

3. Representing primes

Assumptions in parts (iii) and (iv) of Lemma 6 hold whenever $c_n > 1$ for every $n \in \mathbb{N}$. Hence we already know that the set of possible numbers α in Theorem 3 is nowhere dense and has measure zero. In order to show that it is non-empty (and indeed has the cardinality of the continuum), we need a result from the author's paper [4], where she proved that

$$\sum_{\substack{p_{n+1}-p_n > x^{1/2} \\ x \leqq p_n \leqq 2x}} (p_{n+1} - p_n) \ll x^{2/3}, \tag{3}$$

where p_n is the nth prime number.

Actually the proof of this implies the following stronger result (see [4, Lemma 1.2 and its proof] which show how the sum in (3) is attacked).

Lemma 7. *There exist positive constants $d' < 1$ and D' such that, for every sufficiently large x, the interval $[x, 2x]$ contains at most $D'x^{1/6}$ disjoint intervals $\left[n, n + n^{1/2}\right]$ for which*

$$\pi\left(n + n^{1/2}\right) - \pi(n) \leqq \frac{d' n^{1/2}}{\log n}.$$

Remark 8. This can be compared with Lemma 1 which told us that the number of primes in every interval $[x, x + x^{\gamma}]$ is of the expected order of magnitude when $\gamma \geqq 21/40$. Assuming the Riemann hypothesis, the admissible range can be extended to $\gamma \geqq 1/2 + \varepsilon$. Lemma 7 says unconditionally that, for $\gamma = 1/2$ there are very few exceptional intervals. We will later extend this for $\gamma \geqq 1/2$ (Lemma 9 below) and also mention a result for shorter intervals (Lemma 10 below).

We use Lemma 7 to prove Corollary 4. We prove the corollary before turning to Theorem 3, since details are neater in this special case, so the idea can be seen more clearly.

PROOF OF COROLLARY 4. We use Lemma 7 to construct a sequence (a_n) consisting of primes satisfying conditions of Lemma 5 with $\lambda_n(x) = x^2$. A prime sequence (a_n) clearly satisfies the conditions of Lemma 5 if

$$a_0 \geqq 4 \quad \text{and} \quad a_{n+1} \in \left[a_n^2, a_n^2 + a_n\right].$$

Next we construct such a sequence recursively.

Let d' and D' be as in Lemma 7. Let a_0 be a large prime number such that the interval $\mathcal{A}_1 = [a_0^2, a_0^2 + a_0]$ contains at least $d'a_0/(2\log a_0)$ primes. Such a_0 can be found by the prime number theorem.

Now we proceed by induction. Let $k \geqq 0$. We assume that we have chosen prime numbers $a_0, , a_k$ such that each interval

$$\mathcal{A}_{j+1} = \left[a_j^2, a_j^2 + a_j\right], \qquad j = 0, \dots, k$$

contains at least $d'a_j/(2\log a_j)$ primes and

$$a_j \in \mathcal{A}_j \quad \text{for} \quad j = 1, \dots, k.$$

We want to find a prime $a_{k+1} \in \mathcal{A}_{k+1}$ such that the interval $\left[a_{k+1}^2, a_{k+1}^2 + a_{k+1}\right]$ contains at least $d'a_{k+1}/(2\log a_{k+1})$ primes.

Let $\mathcal{P}_{k+1} = \mathcal{A}_{k+1} \cap \mathbb{P}$. For $p \in \mathcal{P}_{k+1}$, the intervals $[p^2, p^2 + p]$ are disjoint and contained in $\left[a_k^4, 2a_k^4\right]$. By Lemma 7 at most $D'a_k^{2/3}$ of them contains less than $d'p/(2\log p)$ primes. But

$$D'a_k^{2/3} < d'a_k/(2\log a_k) \leqq |\mathcal{P}_{k+1}| \tag{4}$$

if a_0 is large enough.

Hence we can choose $a_{k+1} \in \mathcal{P}_{k+1}$ such that the interval $\left[a_{k+1}^2, a_{k+1}^2 + a_{k+1}\right]$ contains at least $d'a_{k+1}/(2\log a_{k+1})$ primes and the induction is finished.

Lemma 5 implies that there exists α such that $a_n = \lfloor \alpha^{2^n} \rfloor$. Since we had multiple choices for a_i at each stage, the set of possible α has the cardinality of the continuum. $\square$

Before proving Theorem 3, we extend Lemma 7 to longer intervals.

LEMMA 9. *There exist positive constants $d < 1$ and D such that, for every sufficiently large x and every $\gamma \in [1/2, 1]$, the interval $[x, 2x]$ contains at most $Dx^{2/3-\gamma}$ disjoint intervals $[n, n + n^\gamma]$ for which*

$$\pi(n + n^\gamma) - \pi(n) \leqq \frac{dn^\gamma}{\log n}. \tag{5}$$

PROOF. Take $d = d'/8$ and $D = 8D'$, where d' and D' are the constants in Lemma 7. Consider a set of $Dx^{2/3-\gamma}$ disjoint intervals $[n, n+n^{\gamma}] \subseteqq [x, 2x]$. They contain at least $4D'x^{1/6}$ disjoint subintervals $[n, n+n^{1/2}]$. By Lemma 7 at least $3D'x^{1/6}$ of these contain more than $d'n^{1/2}/\log n$ primes. Therefore the union of the original intervals contains more than

$$3D'x^{1/6}\frac{d'x^{1/2}}{\log x} = \frac{3d'D'x^{2/3}}{\log x}$$

primes, so at least one of them contains more than

$$\frac{3d'D'x^{\gamma}}{D\log x} > \frac{dn^{\gamma}}{\log n}$$

primes. This implies the claim. $\square$

PROOF OF THEOREM 3. The proof is similar to that of Corollary 4 and so we only sketch it here. We let d and D be as in Lemma 9 and will choose prime numbers $a_0, a_1, a_2, \dots$ such that each interval

$$\mathcal{A}_{j+1} = \left[a_j^{c_{j+1}}, a_j^{c_{j+1}} + a_j^{c_{j+1}-1}\right], \quad j = 0, 1, 2, \dots$$

contains at least $da_j^{c_{j+1}-1}/(c_{j+1}\log a_j)$ primes and $a_j \in \mathcal{A}_j$ for each $j \in \mathbb{N}$.

In the induction step we again let $\mathcal{P}_{k+1} = \mathcal{A}_{k+1} \cap \mathbb{P}$. For $p \in \mathcal{P}_{k+1}$, the intervals

$$\left[p^{c_{k+2}}, p^{c_{k+2}} + p^{c_{k+2}-1}\right]$$

are disjoint and contained in $\left[a_k^{c_{k+1}c_{k+2}}, 2a_k^{c_{k+1}c_{k+2}}\right]$. By Lemma 9 at most

$$Da_k^{c_{k+1}c_{k+2}\left(\frac{2}{3}-\frac{(c_{k+2}-1)}{c_{k+2}}\right)} = Da_k^{c_{k+1}-c_{k+1}c_{k+2}/3}$$

of them contains less than $dp^{c_{k+2}-1}/(c_{k+2}\log p)$ primes. But $c_j \geqq 2$ for every $j \in \mathbb{N}$, so that

$$\text{(6)} \qquad Da_k^{c_{k+1}-c_{k+1}c_{k+2}/3} < \frac{da_k^{c_{k+1}-1}}{c_{k+1}\log a_k} \leqq |\mathcal{P}_{k+1}|$$

if a_0 is large enough. Now the proof can be finished as that of Corollary 4. $\square$

Taking into account the strictness of the inequality (6) (or (4)), it seems that there should be a way to push the method further to get a result with a looser requirement than $c_i \geqq 2$. The reason we are stuck with the bound 2 at the moment is that Lemma 9 and all the work in [4] concerns intervals $\left[n, n+n^{1/2}\right]$.

However there is a companion sum to (3) which deals with shorter intervals. Indeed Peck [7] has shown that

$$\sum_{x \leqq p_n \leqq 2x} (p_{n+1} - p_n)^2 \ll x^{5/4+\varepsilon}. \tag{7}$$

His method gives the following correspondence to Lemma 9.

Lemma 10. *There exist positive constants $d < 1$ and D such that, for every sufficiently large x and every $\gamma \in (0,1]$, the interval $[x, 2x]$ contains at most $Dx^{5/4-2\gamma+\varepsilon}$ disjoint intervals $[n, n+n^{\gamma}]$ for which*

$$\pi(n+n^{\gamma}) - \pi(n) \leqq \frac{c_2 n^{\gamma}}{\log n}.$$

Unfortunately, for intervals of length $\gamma \leqq 1/2$, this is too weak for our purposes. Using this instead of Lemma 9, we would only get Theorem 3 with the requirement $c_i \geqq 2 + \varepsilon$.

Peck [7] used Heath-Brown's identity when he proved Lemma 10. With the same method, he proved the bound (3) with 2/3 replaced by 25/36 (This also appeared in [8].) The current author managed to improve 25/36 to 2/3 in [4] using Harman's sieve method instead of Heath-Brown's identity. It is reasonable to expect that this change of sieve method would also let one improve the exponent in (7) or at least improve Lemma 10 when γ is just below 1/2. This in turn would lead to a looser requirement for c_i.

Assuming the Riemann hypothesis the exponent 5/4 in Lemma 10 can be replaced by 1 by the work of Selberg [9]. This implies that, assuming the Riemann hypothesis, Theorem 3 holds with the requirement $c_i \geqq \left(1+\sqrt{5}\right)/2$.

References

[1] G. Alkauskas and A. Dubickas, Prime and composite numbers as integer parts of powers, *Acta Math. Hungar.,* **105** (2004), 249–256.

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

[3] A. E. Ingham, On the difference between consecutive primes, *Quart. J. Math.,* **8** (1937), 255–266.

[4] K. Matomäki, Large differences between consecutive primes, *Quart. J. Math.,* **58** (2007), 489–518.
[5] W. H. Mills, A prime representing function, *Bull. Amer. Math. Soc.,* **53** (1947), 604.
[6] I. Niven, Functions which represent prime numbers, *Proc. Amer. Math. Soc.,* **2** (1951), 753–755.
[7] A. S. Peck, *On the differences between consecutive primes,* PhD thesis (University of Oxford, 1996).
[8] A. S. Peck, Differences between consecutive primes, *Proc. London Math. Soc.* (3), **76** (1998), 33–69.
[9] A. Selberg, On the normal density of primes in small intervals, and the difference between consecutive primes, *Arch. Math. Naturvid,* **47** (1943), 87–105.
[10] E. M. Wright, A class of representing functions, *J. London Math. Soc.,* **29** (1954), 63–71.