A NOTE ON PRIMES WITH PRIME INDICES

BLAŽEJ ŽMIJA

Abstract. Let \( n, k \in \mathbb{N} \) and let \( p_n \) denote the \( n \)th prime number. We define \( p_n^{(k)} \) recursively as \( p_n^{(1)} := p_n \) and \( p_n^{(k)} = p_{p_n^{(k-1)}} \), that is, \( p_n^{(k)} \) is the \( p_n^{(k-1)} \)th prime.

In this note we give answers to some questions and prove a conjecture posed by Miska and Tóth in their recent paper concerning subsequences of the sequence of prime numbers. In particular, we establish explicit upper and lower bounds for \( p_n^{(k)} \). We also study the behaviour of the counting functions of the sequences \( (p_n^{(k)})_{k=1}^{\infty} \) and \( (p_k^{(k)})_{k=1}^{\infty} \).

1. Introduction

Let \( (p_n)_{n=1}^{\infty} \) be the sequence of consecutive prime numbers. In a recent paper [3] Miska and Tóth introduced the following subsequences of the sequence of prime numbers: \( p_n^{(1)} := p_n \) and for \( k \geq 2 \)

\[
p_n^{(k)} := p_{p_n^{(k-1)}},
\]

In other words, \( p_n^{(k)} \) is the \( p_n^{(k-1)} \)th prime. They also defined

\[
\Diag^\mathbb{P} := \{ p_k^{(k)} | k \in \mathbb{N} \},
\]

\[
\mathbb{P}_n^T := \{ p_n^{(k)} | k \in \mathbb{N} \}
\]

for each positive integer \( n \).

The main motivation in [3] was the known result that the set of prime numbers is \((R)\)-dense, that is, the set \( \{ \frac{p}{q} | p, q \in \mathbb{P} \} \) is dense in \( \mathbb{R}_+ \) (with respect to the natural topology on \( \mathbb{R}_+ \)). It was proved in [3] that for each \( k \in \mathbb{N} \) the sequence \( \mathbb{P}_k := (p_n^{(k)})_{n=1}^{\infty} \) is \((R)\)-dense. This result might be surprising, because the sequences \( \mathbb{P}_k \) are very sparse. In fact, for each \( k \) set \( \mathbb{P}_k \) is a zero asymptotic density subset of \( \mathbb{P}_k \). On the other hand, it was showed, that the sequences \( (p_n^{(k)})_{k=1}^{\infty} \) for each fixed \( n \in \mathbb{N} \), and \( (p_k^{(k)})_{k=1}^{\infty} \) are not \((R)\)-dense.

Results of another type that were proved in [3] concern the asymptotic behaviour of \( p_n^{(k)} \) as \( n \to \infty \), or as \( k \to \infty \). In particular, as \( n \to \infty \), we have for each \( k \in \mathbb{N} \)

\[
p_n^{(k)} \sim n(\log n)^k,
\]

\[
p_{n+1}^{(k)} \sim p_n^{(k)},
\]

\[
\log p_n^{(k)} \sim \log n
\]

by [3] Theorem 1]. Some results from [3] concerning \( p_n^{(k)} \) as \( k \to \infty \) are mentioned later.

For a set \( A \subseteq \mathbb{N} \) let \( A(x) \) be its counting function, that is,

\[
A(x) := \#(A \cap [1, x]).
\]

Miska and Tóth posed four questions concerning the numbers \( p_n^{(k)} \):

A. Is it true that \( p_n^{(k)} \sim p_k^{(k)} \) as \( k \to \infty \)?
B. Are there real constants \( c > 0 \) and \( \beta \) such that

\[
\exp \mathbb{P}_n^T(x) \sim cx(\log x)^\beta
\]

for each \( n \in \mathbb{N} \)?
C. Are there real constants \( c > 0 \) and \( \beta \) such that

\[
\exp \Diag^\mathbb{P}(x) \sim cx(\log x)^\beta
\]

for each \( n \in \mathbb{N} \)?
D. Is it true that

\[
\Diag^\mathbb{P}(x) \sim \mathbb{P}_n^T(x)
\]

for each \( n \in \mathbb{N} \)?

2010 Mathematics Subject Classification. 11A41.
Key words and phrases. prime number, subsequence.

During the preparation of the work, the author was a scholarship holder of the Kartezjusz program funded by the Polish National Center for Research and Development.

1
The aim of this paper is to give answers to question B, C and D. The main ingredients of our proofs are the following inequalities:

\[ n \log n < p_n < 2n \log n. \] (1)

The first inequality holds for all \( n \geq 2 \), and the second one for all \( n \geq 3 \). For the proofs, see [4]. In Section 2 we use (1) in order to show explicit bounds for \( p_k^{(k)} \). In particular, for all \( n > e^{4200} \) we have:

\[
\log p_n^{(k)} = k(\log k + \log \log k + O_n(1)),
\]

\[
\log p_k^{(k)} = k(\log k + \log \log k + O(\log \log \log k)),
\]

as \( k \to \infty \), where the implied constant in the first line may depend on \( n \), see Theorem 3.1 below. In consequence, we improve the (in)equalities

\[
\lim_{k \to \infty} \frac{p_n^{(k)}}{k \log k} = 1,
\]

\[
1 \leq \lim \inf_{k \to \infty} \frac{p_k^{(k)}}{k \log k} \leq \lim \sup_{k \to \infty} \frac{p_k^{(k)}}{k \log k} \leq 2.
\]

that appeared in [3]. Then we show in Section 3 that the answers to questions B and C are negative (Corollary 3.4), while the one for question D is affirmative (Theorem 3.2). In fact, we find the following relation:

\[
\mathbb{P}_n^T(x) \sim \text{Diag P}(x) \sim \frac{\log x}{\log \log x}
\]

for all positive integers \( n \).

In their paper, Miska and Tóth also posed a conjecture, that we state here as a proposition, since it is in fact a consequence of a result that had already appeared in [3].

**Proposition 1.1.** Let \( n \in \mathbb{N} \) be fixed. Then

\[
\frac{p_n^{(k)}}{p_k^{(k)}} \to 0
\]

as \( k \to \infty \).

**Proof.** Let \( k > p_n \). Then

\[
0 \leq \frac{p_n^{(k)}}{p_k^{(k)}} < \frac{p_n^{(k)}}{p_{p_n}^{(k)}} = \frac{p_n^{(k)}}{p_n^{(k+1)}},
\]

The expression on the right goes to zero as \( k \) goes to infinity, as was proved in [3] Corollary 3].

It is worth to note, that primes with prime indices have already appeared in the literature, for example in [1] and [2]. However, according to our best knowledge, our paper is the second one (after [3]), where the number of iterations of indices, that is, the number \( k \) in \( p_n^{(k)} \), is not fixed.

Throughout the paper we use the following notation: \( \log x \) denotes the natural logarithm of \( x \), and for functions \( f \) and \( g \) we write \( f \sim g \) if \( \lim_{x \to \infty} \frac{f(x)}{g(x)} = 1 \), \( f = O(g) \) if there exists a positive constant \( c \) such that \( f(x) < cg(x) \), and \( f = o(g) \) if \( \lim_{x \to \infty} \frac{f(x)}{g(x)} = 0 \).

## 2. Upper and lower bounds for \( p_n^{(k)} \)

In this Section, we find explicit upper and lower bounds for \( p_n^{(k)} \). We start with the upper bound.

**Lemma 2.1.** Let \( n \geq 9 \). Then for each \( k \in \mathbb{N} \) we have:

\[
p_n^{(k)} < 2^{2k-1} \cdot n \cdot (k-1)! \cdot \left( \log(\max\{k,n\}) \right)^k.
\]

In particular,

\[
p_n^{(k)} < (4 \cdot k \log k)^k
\]

for \( k \geq n \).
where the last inequality follows from the well-known inequality \( y > \frac{1}{2} \sqrt{x} \).

At first, we prove that functions \( g(x) \) is increasing. Indeed, if 
\[
 f(x) := \log \left( \frac{x}{x+1} \right)^{x+1} = (x+1)(\log x - \log(x+1)),
\]
then 
\[
 f'(x) = \log x - \log(x+1) + (x+1) \left( \frac{1}{x} - \frac{1}{x+1} \right) = \frac{1}{x} - \log \left( \frac{1+1}{x} \right) > 0,
\]
where the last inequality follows from the well-known inequality \( y > \log(1+y) \) used with \( y = \frac{1}{x} \). Hence, we can bound 
\[
 \left( \frac{x}{x+1} \right)^{x+1} \leq \left( \frac{4200}{4201} \right)^{x+1}
\]
for all \( x \geq 4200 \).

Now we need to find a lower bound for \( \left( \frac{\log x}{\log(x+1)} \right)^{x+1} \). Let us write 
\[
 \left( \frac{\log x}{\log(x+1)} \right)^{x+1} = \left[ \left( 1 - \frac{1}{\log(x+1) - \log x} \right)^{\log(x+1) - \log x} \right]^{\frac{1}{\log(x+1) - \log x}} \left( 1 + \frac{1}{x} \right) \left( 1 + \frac{1}{x} \right)^{\frac{1}{\log x - \log(x+1)}} (1 + \frac{1}{x})
\]
At first, we prove that functions \( g(t) := (1 - \frac{1}{t})^t \) and \( h(x) := \frac{\log(x+1)}{\log(x+1) - \log x} \) are increasing. For the function \( g(t) \) it is enough to observe, that \( \log g(t) = f(t-1) \) and the function \( f(x) \) is increasing. For the function \( h(x) \) we have: 
\[
 h'(x) = \frac{1}{(\log(x+1) - \log x)^2} [\log(x+1) - \log x - \log(x+1) \left( \frac{1}{x+1} - \frac{1}{x} \right)] 
\]
\[
 = \frac{1}{(\log(x+1) - \log x)^2} \left[ \frac{\log(x+1) - \log x}{x+1} - \frac{\log x}{x+1} \right] > 0.
\]
The fact that the functions \( g(t) \) and \( h(x) \) are increasing, together with the properties \( g(h(4200)) \in (0,1) \) and \( \log \left( 1 + \frac{1}{x} \right)^x < 1 \), give us 
\[
 \left( \frac{\log x}{\log(x+1)} \right)^{x+1} \geq \left[ g(h(4200)) \right]^{\left( \frac{1}{\log(x+1) - \log x} \right) \left( \frac{1}{x+1} \right) \left( 1 + \frac{1}{x} \right)} > \left[ g(h(4200)) \right]^{\frac{1}{\log x - \log(x+1)}} \left( \frac{4200}{\log 4201} \right)^{\frac{4200}{\log 4201} - \frac{1}{\log 4201}}
\]
for all \( x \geq 4200 \).

\[\boxed{\text{Lemma 2.2. Let } L(x) := \left( \frac{x}{x+1} \right)^{x+1} \left( \frac{\log x}{\log(x+1)} \right)^{x+1} \text{.}
\]
Then we have 
\[ L(x) > 0.32627 \]
for all \( x \geq 4200 \).

Proof. We proceed by induction on \( k \). For \( k = 1 \) it is a simple consequence of (1). Then the second induction step goes as follows: let us denote \( m := \max\{n, k\} \). Observe that \( (k-1)! < (k-1)^{k-1} < m^{k-1} \) and \( 4 \log m < m \) for \( m \geq 9 \). Hence, 
\[
 p_{n}^{(k+1)} \leq p_{n}^{k+1} \log p_{n}^{(k)} < 2 \cdot 2^{k-1} \cdot n \cdot (k-1)! \cdot \log [2^{k-1} \cdot n \cdot (k-1)! \cdot \log m] 
\]
\[
 < 2^{k} \cdot n \cdot (k-1)! \cdot \log [4 \cdot m \cdot m^{k-1} \cdot (log m)] = 2^{k} \cdot n \cdot (k-1)! \cdot \log m^{k} \log [4 \cdot m \cdot m^{k}] 
\]
\[
 \leq 2^{k} \cdot n \cdot k! \cdot (\log m)^{k} \cdot \log |m|^{2} \leq 2^{k} \cdot n \cdot k! \cdot (\log m)^{k+1}.
\]
The second part of the statement is an easy consequence of the first part and the inequalities \((k-1)! < k^{k-1}\) and \( n \leq k \).
Therefore, if \( N \) show the following inequalities:

\[
L(x) \geq \left( \frac{4200}{4201} \right)^{4201} \left( \frac{\log 4200}{\log 4201} \right)^{4201} \approx 0.3262768 > 0.32627.
\]

The proof is finished. \( \square \)

In the next lemma we provide a lower bound for \( p_n^{(k)} \).

**Lemma 2.3.** If \( n > e^{4200} \), then for all \( k \geq \lfloor \log n \rfloor \) we have

\[
p_n^{(k)} > n^{(k)}.
\]

**Proof.** First, let us observe that a simple induction argument on \( k \) implies the inequality

\[
(4)
\]

Indeed, for \( k = 1 \) this follows from left inequality in \( (1) \). Using the same inequality we get also

\[
p_n^{(k+1)} > p_n^{(k)} \log p_n^{(k)} > n^{(k)} \log(n^{(k)}) > n^{(k+1)}
\]

and hence \( (4) \).

Now we show that the inequality from the statement is true for \( k = \lfloor \log n \rfloor \). Because of \( (4) \) it is enough to show:

\[
n^{(k)} > \left( \frac{e \cdot k \log k}{\log \log n} \right)^k,
\]

or equivalently, after taking logarithms we get

\[
\log n + \lfloor \log n \rfloor \log \log n > \lfloor \log n \rfloor + \lfloor \log n \rfloor \log \lfloor \log n \rfloor + \lfloor \log n \rfloor \log \lfloor \log n \rfloor - \lfloor \log n \rfloor \log \log \log n.
\]

This is equivalent to the inequality

\[
(\log n - \lfloor \log n \rfloor) + \lfloor \log n \rfloor (\log n - \log \lfloor \log n \rfloor) + \lfloor \log n \rfloor (\log \log n - \log \lfloor \log n \rfloor) > 0,
\]

which is obviously true.

In order to finish the proof, we again use the induction argument. The inequality from the statement of our lemma is true for \( k = \lfloor \log n \rfloor \). Assume it holds for some \( k \geq \lfloor \log n \rfloor \). Then by \( (4) \) and the induction hypothesis we get

\[
p_n^{(k+1)} > p_n^{(k)} \log p_n^{(k)} > \left( \frac{e \cdot k \log k}{\log \log n} \right)^k \log \left( \frac{e \cdot k \log k}{\log \log n} \right)^k.
\]

It is enough to show that for all \( n > e^{4200} \) and all \( k \geq \lfloor \log n \rfloor \) we have

\[
\left( \frac{e \cdot k \log k}{\log \log n} \right)^k \log \left( \frac{e \cdot k \log k}{\log \log n} \right)^k > \left( \frac{e \cdot (k+1) \log (k+1)}{\log \log n} \right)^{k+1}.
\]

This is equivalent to

\[
k^{k+1} (\log k)^k \left[ \log k + \log \left( \frac{e \cdot k \log k}{\log \log n} \right) \right] > \frac{e}{\log \log n} (k+1)^{k+1} (\log (k+1))^{k+1}.
\]

Recall, that we assume that \( k \geq \lfloor \log n \rfloor \). Thus \( e \log k > \log \log n \). Therefore, it is enough to show the following inequalities:

\[
k^{k+1} (\log k)^k > \frac{e}{\log \log n} (k+1)^{k+1} (\log (k+1))^{k+1},
\]

or equivalently

\[
\left( \frac{k}{k+1} \right)^{k+1} \left( \frac{\log k}{\log (k+1)} \right)^{k+1} > \frac{e}{\log \log n}.
\]

Notice that the left-hand side expression of the last inequality is equal to \( L(k) \), where the function \( L(x) \) is defined in the statement of Lemma 2.2. If \( n > e^{4200} \), then \( k \geq \lfloor \log n \rfloor \geq 4200 \), and Lemma 2.2 implies \( L(k) > 0.32627 \).

Therefore, if \( N := \max \{ \lfloor e^{4200} \rfloor, \lfloor e^{e^{0.32627}} \rfloor \} = \lfloor e^{4200} \rfloor \), then for all \( n > N \) and \( k \geq \lfloor \log n \rfloor \) we have:

\[
L(k) > 0.32627 \geq \frac{e}{\log \log N} > \frac{e}{\log \log n}.
\]

This finishes the proof. \( \square \)
3. Main results

We begin this section by a theorem that provides good information about asymptotic growth of \( \log p_n^{(k)} \) for large fixed \( n \), and for \( \log p_k^{(k)} \) as \( k \to \infty \).

**Theorem 3.1.**

1. Let \( n > e^{4200} \). Then
   \[
   \log p_n^{(k)} = k(\log k + \log \log k + O_n(1))
   \]
as \( k \to \infty \), where the implied constant may depend on \( n \).

2. We have
   \[
   \log p_k^{(k)} = k(\log k + \log \log k + O(\log \log \log k))
   \]
as \( k \to \infty \).

**Proof.** If \( n > e^{4200} \) and \( k \geq n \), Lemmas 2.1 and 2.3 give us:
\[
\left( \frac{e \cdot k \log k}{\log \log n} \right)^k < p_n^{(k)} < (4 \cdot k \log k)^k.
\]
After taking logarithms, we simply get the first part of our theorem. In order to get the second part, we need to put \( n = k \) and repeat the reasoning. \( \square \)

Now we give the answer to Question D.

**Theorem 3.2.** For each \( n \in \mathbb{N} \) we have
\[
\operatorname{Diag} \mathbb{P}(x) \sim \mathbb{P}_n^{T}(x)
\]
as \( x \to \infty \).

**Proof.** From [3, Theorem 6] we know that
\[
\mathbb{P}_m^{T}(x) \sim \mathbb{P}_n^{T}(x)
\]
for each \( m, n \in \mathbb{N} \). Therefore, it is enough to prove \( \operatorname{Diag} \mathbb{P}(x) \sim \mathbb{P}_n^{T}(x) \) for some sufficiently large \( n \).
Let \( n = \lfloor e^{4200} \rfloor + 100 \). We use the idea from the proof of [3, Theorem 17]. Let \( k \) be a large real number. Let \( k \) be such that \( p_k^{(k)} \leq x < p_{k+1}^{(k+1)} \). Then \( \operatorname{Diag} \mathbb{P}(x) = k \). By [3, Theorem 8] and Theorem 3.1 above we have
\[
\frac{\mathbb{P}_n^{T}(x)}{\operatorname{Diag} \mathbb{P}(x)} \leq 1 + \frac{\log p_{k+1}^{(k+1)} - \log p_k^{(k)}}{k \log \log p_k^{(k)}} - \frac{(1 + o(1))(k + 1) \log(k + 1)}{k \log [1 + o(1)k \log k]} - \frac{(1 + o(1))k \log k}{k \log [(1 + o(1))k \log k]}
\]
\[
= 1 + \frac{1}{k} \log(k + 1) \log k + \log [1 + o(1) \log k] - (1 + o(1)) \frac{\log k \log(k + 1)}{\log k + \log [(1 + o(1)) \log k]}
\]
The whole last expression goes to 1 as \( k \) goes to infinity. On the other hand, \( \operatorname{Diag} \mathbb{P}(x) \leq \mathbb{P}_n^{T}(x) \) for \( x \geq p_n^{(n)} \) and we get the result. \( \square \)

The answers to Questions B and C will follow from our next result, which is of independent interest.

**Theorem 3.3.**

1. Let \( n \in \mathbb{N} \). Then
   \[
   \mathbb{P}_n^{T}(x) \sim \frac{\log x}{\log \log x}.
   \]

2. We have
   \[
   \operatorname{Diag} \mathbb{P}(x) \sim \frac{\log x}{\log \log x}.
   \]

**Proof.** In view of Theorem 3.2, it is enough to show the statement for the function \( \operatorname{Diag} \mathbb{P}(x) \). Let us fix an arbitrarily small number \( \varepsilon > 0 \) and take a sufficiently large real number \( x \) and find \( k \) such that \( p_k^{(k)} \leq x < p_{k+1}^{(k+1)} \). Then \( \operatorname{Diag} \mathbb{P}(x) = k \) and by Lemmas 2.1 and 2.3 we have
\[
k^k < x < k^{(1 + \varepsilon)}k.
\]
Let us write \( x = e^y \). Then
\[
k \log k \leq y < (1 + \varepsilon)k \log k.
\]
If \( y \) is sufficiently large, this implies
\[
(1 - \varepsilon) \frac{y}{\log y} < k < (1 + \varepsilon) \frac{y}{\log y}.
\]

Indeed, if \( k \leq (1 - \varepsilon) \frac{y}{\log y} \), then
\[
y < (1 + \varepsilon) k \log k \leq (1 - \varepsilon^2) \frac{y}{\log y} \log \left( 1 + \frac{y}{\log y} \right) < (1 - \varepsilon^2) \left( 1 - \frac{\log \log y}{\log y} \right) y,
\]
which is impossible. Similarly, if \( k \geq (1 + \varepsilon) \frac{y}{\log y} \), then
\[
y > k \log k \geq (1 + \varepsilon) \frac{y}{\log y} \log \left( 1 + \frac{y}{\log y} \right) > (1 + \varepsilon) \left( 1 - \frac{\log \log y}{\log y} \right) y.
\]
The above inequality cannot hold if \( y \) is sufficiently large.

If we go back to \( k = \text{Diag} \mathbb{P}(x) \) and \( y = \log x \) in (5), we get
\[
(1 - \varepsilon) \frac{\log x}{\log \log x} < \text{Diag} \mathbb{P}(x) < (1 + \varepsilon) \frac{\log x}{\log \log x}.
\]
The number \( \varepsilon > 0 \) was arbitrary, so the result follows. \( \square \)

Corollary 3.4. There do not exist constants \( c > 0 \) and \( \beta \) such that
\[
\exp \mathbb{P}^T_n(x) \sim cx(\log x)^\beta
\]
for some \( n \), or
\[
\exp \text{Diag} \mathbb{P}(x) \sim cx(\log x)^\beta.
\]

Proof. We prove the result only for \( \text{Diag} \mathbb{P}(x) \). The case of \( \mathbb{P}_T(x) \) is analogous.
Assume to the contrary, that \( \exp \text{Diag} \mathbb{P}(x) \sim cx(\log x)^\beta \) for some \( c > 0 \) and \( \beta \). Then \( \exp \text{Diag} \mathbb{P}(x) = (1 + o(1))cx(\log x)^\beta \). This, after taking logarithms on both sides, implies \( \text{Diag} \mathbb{P}(x) = \log x + O(\log \log x) \), contradicting Theorem 3.3. \( \square \)

Acknowledgement

I would like to express my gratitude to Piotr Miska, who informed me about the problem and helped to improve the quality of the paper. I would also like to thank Carlo Sanna, whose comments allowed me to simplify a part of the paper.

References

[1] K. Broughan, R. Barnett, On the subsequence of primes having prime subscripts, Journal of Integer Sequences 12, article 09.2.3, 2019.
[2] J. Bayless, D. Klyve, T. Oliveira e Silva, New bounds and computations on prime-indexed primes, Integers 13: A43:1-A43:21, 2013.
[3] P. Miska, J. Tóth, On interesting subsequences of the sequence of primes, preprint: arXiv:1908.10421.
[4] J. B. Rosser, L. Schoenfeld, Approximate formulas for some functions of prime numbers, Illinois J. Math. 6, no. 1 , 64–94, 1962. doi:10.1215/ijm/1255631807.