On the Asymptotic Density of Prime k-tuples and a Conjecture of Hardy and Littlewood

L. Tóth

Grand Duchy of Luxembourg
Rue des Tanneurs 7
L-6790 Grevenmacher
E-mail: uk.laszlo.toth@gmail.com

Abstract: In 1922 Hardy and Littlewood proposed a conjecture on the asymptotic density of admissible prime k-tuples. In 2011, Wolf computed the "Skewes number" for twin primes, i.e., the first prime at which a reversal of the Hardy-Littlewood inequality occurs. In this paper, we find "Skewes numbers" for 8 more prime k-tuples and provide numerical data in support of the Hardy-Littlewood conjecture. Moreover, we present several algorithms to compute such numbers.
Key words: prime k-tuple, asymptotic density, conjecture, Skewes number

I. INTRODUCTION

Let $(a_1, a_2, \ldots, a_k)$ denote a monotonically increasing sequence of positive even integers and let p be a prime number. Then if the numbers $p + a_i$ for all $1 \le i \le k$ are prime, the sequence $P = (p, p + a_1, p + a_2, \ldots, p + a_k)$ is a *prime k-tuple*. Moreover, if these numbers do not form a complete residue class with respect to any prime, P is *admissible*. Hardy and Littlewood [5] made several conjectures concerning the infinitude of admissible prime k-tuples. In particular, they conjectured that their asymptotic density can be calculated in terms of the $(a_1, a_2, \ldots, a_k)$ as follows.

Conjecture 1 (Asymptotic density of prime k-tuples) *Let $P = (p, p + a_1, p + a_2, \ldots, p + a_k)$ denote an admissible prime k-tuple and let $\pi_P(n)$ denote the number of primes p less than a positive integer n such that for all $1 \le i \le k$, $p + a_i$ is prime. Then*

$$\pi_P(n) \sim C_{a_1, a_2, \ldots, a_k} \int_2^n \frac{dt}{\log^{k+1} t},$$

where $C_{a_1, a_2, \ldots, a_k}$ is a constant obtained through a product over all primes q greater than 2 and the amount of

distinct residues of $a_1, a_2, ..., a_k$ modulo q, denoted by $w(q; a_1, a_2, ..., a_k)$, as follows:

$$C_{a_1, a_2, \ldots, a_k} = 2^k \prod_q \frac{1 - \frac{w(q; a_1, a_2, \ldots, a_k)}{q}}{(1 - \frac{1}{q})^{k+1}}.$$

For example, when $k = 1$ and $a_1 = 2$ (i.e., the twin primes), we have $C_2 = 2 \prod_q \left(1 - \frac{1}{(q-1)^2}\right) = 1.320323632\ldots$, called the twin prime constant (although some authors prefer defining $\frac{C_2}{2}$ as the twin prime constant).

If the density described in Conjecture 1 is true, it immediately implies the infinitude of the corresponding prime tuple, so any numerical data in its favor is of value. There are several ways to study this density, for instance by analyzing the number of sign changes in the difference

$$\delta_P(n) = \pi_P(n) - C_{a_1, a_2, \ldots, a_k} \int_2^n \frac{dt}{\log^{k+1} t}. \quad (1)$$

This is comparable to the study of the so-called Skewes number, the first positive integer n such that $\pi(n) > \text{Li}(n)$, where $\text{Li}(n)$ denotes the logarithmic integral $\int_2^n \frac{dt}{\log t}$ and $\pi(n)$ is the usual prime-counting function. The existence of such a number was shown first by Littlewood [6] and an upper

bound was given by Skewes ([7] assuming the Riemann Hypothesis, [8] without assuming it). In this paper, we extend this definition by formally defining the "Skewes number" for prime k-tuples as follows.

Definition 1 (Skewes number for prime k-tuples)
Let p denote a prime number, k a positive nonzero integer and let a_i for all $1 \leq i \leq k$ be positive integers. Furthermore, let $P = (p, p + a_1, p + a_2, \ldots, p + a_k)$ denote an admissible prime k-tuple, $\pi_P(n)$ the counting function for P and $C_{a_1,a_2,\ldots,a_k}$ the associated prime tuple constant. Then the first prime $p \in P$ that violates

$$\pi_P(n) < C_{a_1,a_2,\ldots,a_k} \int_2^n \frac{dt}{\log^{k+1} t},$$

if such a prime exists, is the Skewes number *for P.*

The Skewes number for twin primes was already studied by several authors, including Brent [1] and Wolf [2], which are particularly interesting. First, define $P_2 = (p, p + 2)$ and let $\pi_2(n)$ denote the number of primes not exceeding n such that $n + 2$ is also prime. Furthermore, define $\mathrm{Li}_2(n) = \int_2^n \frac{dt}{\log^2 t}$. Wolf [2] computed the number of sign changes in $\delta_{P_2} = \pi_2(n) - C_2\mathrm{Li}_2(n)$ and found that there are exactly 477118 up to $2^{48} \approx 2.81 \times 10^{14}$. Moreover, the author found the Skewes number for twin primes, 1369391. Based on numerical data, Wolf then proposed the following conjecture.

Conjecture 2 (Wolf) *The number of sign changes in $d_2(n) = \pi_2(n) - C_2\mathrm{Li}_2(n)$ within the interval $n \in (1, T)$ is given by $\frac{\sqrt{T}}{\log T}$.*

Note that this result was obtained after a number of "trials" and no heuristic argument is given in support of the conjecture. The author's results continue the work of Brent [1], who also studied the difference $\delta_{P_2}(n)$ and examined its maxima and minima in various intervals $[a, b]$ up to 8×10^{10} by defining the functions:

$$R_3(a, b) = \max_{p \in P_2 \cap [a,b]} \delta_{P_2}(p)$$

and

$$\rho_3(a, b) = \min_{p \in P_2 \cap [a,b]} \delta_{P_2}(p)$$

(note that we adapted Brent's notation to ours in the above definitions). The author also used the numerical data he obtained in order to give an approximation to Brun's constant. Nicely [4] considered three additional k-tuples: $(p, p+2, p+6)$, $(p, p+4, p+6)$ and $(p, p+2, p+6, p+8)$, and computed the corresponding coefficients in their conjectured asymptotic density (denoted by C in Conjecture 1) to a high

precision. He did not compute the Skewes number for these tuples. Moreover, none of the authors mentioned above provided computer code.

In another paper, motivated by the fact Conjecture 1 implies that some prime tuples have equal asymptotic density, Wolf [3] investigated the asymptotic densities of the twin and cousin primes $(p, p + 4)$. Indeed, we expect that

$$\pi_{\mathrm{twin}}(n), \pi_{\mathrm{cousin}}(n) \sim 2 \prod_{p \geq 3} \frac{p(p-2)}{(p-1)^2} \int_2^n \frac{dt}{\log^2 t}. \quad (2)$$

Wolf looked at the relationship between these two densities by defining the function $y(x)$ as the difference between the number of twin and cousin primes up to a positive integer x. Along with a number of numerical computations, Wolf studied the fractal properties of $y(x)$ by performing a random walk in which $+1$ corresponds to a pair of cousin primes and -1 to a pair of twin primes encountered by the walker. In particular, Wolf observed that $y(x) = 0$ for 2823290 primes up to $x = 2^{43} \approx 8.8 \times 10^{12}$ and noted that the clusters of x where $y(x) = 0$ formed a self-similar set.

I. 1. Scope of this paper

The aim of this paper is threefold; first, to compute the Skewes number for the prime tuples listed below in Tab. 1, second, to provide a flexible algorithm for studying and gathering data on the asymptotic density of prime k-tuples and third, to provide computational data in support of the first Hardy-Littlewood conjecture.

In the remainder of this paper we shall denote by $\mathrm{Li}_k(n)$ the integral $\int_2^n \frac{dt}{\log^k t}$ and by $\pi_P(n)$ the prime-counting function related to the prime tuple P. Tab. 1 shows the admissible prime k-tuples we consider in this paper along with their conjectured asymptotic densities. Note that we include the twin primes within our study for the sake of completeness and in order to confirm Wolf's result that the corresponding Skewes number is indeed 1369391.

II. COMPUTATIONAL AND ALGORITHMIC METHODS

In this section we present the computational and algorithmic methods we used to obtain our results in Section III. These include finding the Skewes number for the prime k-tuples listed in Tab. 1 as well as gathering other relevant numerical data in support of Conjecture 1.

Note that the pseudo-code presented in this section, supplemented with code used for plotting purposes and gathering other secondary data was implemented in Wolfram Mathematica 11.1 and executed on an Intel Core i7-7800X CPU @3.50 GHz and 16 GB RAM to produce the results in Section III.

Tab. 1. Prime k-tuples considered in this paper

Tuple	Definition	Conjectured asymptotic density
P_{2a}	$(p, p+2)$	$2 \prod_{p \geq 3} \frac{p(p-2)}{(p-1)^2} \mathrm{Li}_2(n)$
P_{2b}	$(p, p+4)$	$2 \prod_{p \geq 3} \frac{p(p-2)}{(p-1)^2} \mathrm{Li}_2(n)$
P_{3a}	$(p, p+2, p+6)$	$\frac{9}{2} \prod_{p \geq 5} \frac{p^2(p-3)}{(p-1)^3} \mathrm{Li}_3(n)$
P_{3b}	$(p, p+4, p+6)$	$\frac{9}{2} \prod_{p \geq 5} \frac{p^2(p-3)}{(p-1)^3} \mathrm{Li}_3(n)$
P_{4a}	$(p, p+2, p+6, p+8)$	$\frac{27}{2} \prod_{p \geq 5} \frac{p^3(p-4)}{(p-1)^4} \mathrm{Li}_4(n)$
P_{4b}	$(p, p+4, p+6, p+10)$	$27 \prod_{p \geq 5} \frac{p^3(p-4)}{(p-1)^4} \mathrm{Li}_4(n)$
P_{5a}	$(p, p+2, p+6, p+8, p+12)$	$\frac{15^4}{2^{11}} \prod_{p \geq 7} \frac{p^4(p-5)}{(p-1)^5} \mathrm{Li}_5(n)$
P_{5b}	$(p, p+4, p+6, p+10, p+12)$	$\frac{15^4}{2^{11}} \prod_{p \geq 7} \frac{p^4(p-5)}{(p-1)^5} \mathrm{Li}_5(n)$
P_6	$(p, p+4, p+6, p+10, p+12, p+16)$	$\frac{15^5}{2^{13}} \prod_{p \geq 7} \frac{p^5(p-6)}{(p-1)^6} \mathrm{Li}_6(n)$

The core of our algorithm relies on computing the logarithmic integral within successive intervals, the upper and lower bounds corresponding to consecutive primes of a given k-tuple. This method was also employed by Wolf [2] to compute the Skewes number for twin primes. Unfortunately, he did not give any code in support of his algorithm or consider any special cases that might arise, such as insufficient computing memory or other extensions that allow recording additional data. Here we present two algorithms as pseudo-code. The first outlines our basic methodology while the second considers limitations in computing resources. Both of these algorithms are flexible, i.e., they can be easily modified to record intermediary data and fine-tune the parameters.

We will now follow with a short presentation of the main algorithm. We begin by defining our inputs. These are:

- An array of primes in the prime k-tuple **P** in the range $[2, n]$,
- The corresponding prime tuple constant **C**, shown in Tab. 1,
- The corresponding logarithmic integral function **LiP**[].

The main program loop then computes the logarithmic integral within successive intervals and stores the cumulated value in a variable. This value is then multiplied by **C** and compared to the current value of π_P, which conveniently is the index of the current prime in **P**. The program contin-

ues until a reversal of the sign in the difference between π_P and **C** times the cumulated logarithmic interval is reached. At this point, the Skewes number is returned. Algorithm 1 illustrates this method. Please note that the first prime is denoted by p_1 in the algorithm below.

Algorithm 1 Finding the Skewes number for a prime k-tuple

Input: Prime k-tuple **P** within the range $[2, n]$, prime tuple constant **C**, **LiP**[ ] function
Output: Skewes number **S**
$cumulLogInt \leftarrow 0$
$currentSign \leftarrow \mathrm{Sign}[-1]$
for all $p_i \in$ **P do**
    if $i = 1$ **then**
        $intLBound \leftarrow 2$
    else
        $intLBound \leftarrow p_{i-1}$
    end if
    $intUBound \leftarrow p_i$
    $cumulLogInt \leftarrow \mathbf{LiP}[intLBound, intUBound]$
    $diff \leftarrow (i - \mathbf{C} \times cumulLogInt)$
    $newSign \leftarrow \mathrm{Sign}[diff]$
    if $currentSign \neq newSign$ **then**
        return p_i
    end if
end for

A number of non-essential features were omitted from Algorithm 1. For instance, it is desirable to record some of the intermediary values computed within the main loop such as the differences between $\pi_P(i)$ and its conjectured density. For instance, the last **if** clause can be enlarged with a block of code allowing to count the number of sign changes instead of returning the Skewes number. Our Mathematica implementation of the algorithm does just that, and some plots showing this data are presented in Section III.

Obviously the biggest challenge in Algorithm 1 lies in providing the input to the algorithm, i.e., a list of primes within a given prime k-tuple up to a very high numerical limit. We note that the advantage of pre-computing such a list of primes instead of iterating through integers and testing for primality of other potential members of the tuple lies in the fact that highly efficient algorithms already exist within most mathematical software packages that are able to yield such an array in a computationally short amount of time. For instance, Wolfram Mathematica has the following one-line solution (taking the example of cousin primes):

```
cousinPrimes=Select[Prime[Range[PrimePi[
upperLimit]]], PrimeQ[# + 4] &];
```

with `upperLimit` replaced by any numerical upper bound. Of course, the average computer will quickly begin to struggle with the above code as larger values of `upperLimit`

require generous amounts of memory to hold all the primes. For this reason, we present a modified version of Algorithm 1 such that, when a large enough upper bound is reached, we discard the contents of the current prime array and fill it with another "chunk" of equally large size; the current lower bound thus becomes the previous upper bound. We then use an offset in order to compensate for the prime counts in previous chunks. This revised version of the algorithm is presented in Algorithm 2.

Algorithm 2 Finding the Skewes number for a prime k-tuple, taking into account memory limitations

---

Input: chunkSize: the number of primes within a chunk, prime tuple constant **C**, **LiP**[ ] function
Output: Skewes number **S**
$offset \leftarrow 0$
for $chunk \leftarrow 2$, **increment chunkSize, do**
    P $\leftarrow$ prime k-tuple within the range $[chunk, chunk+\textbf{chunkSize}]$
    $cumulLogInt \leftarrow 0$
    $currentSign \leftarrow \text{Sign}[-1]$
    for all $p_i \in \textbf{P}$ **do**
        if $i = 1$ **then**
            $intLBound \leftarrow 2$
        else
            $intLBound \leftarrow p_{i-1}$
        end if
        $intUBound \leftarrow p_i$
        $cumulLogInt \leftarrow \textbf{LiP}[intLBound, intUBound]$
        $diff \leftarrow (i + offset - \textbf{C} \times cumulLogInt)$
        $newSign \leftarrow \text{Sign}[diff]$
        if $currentSign \neq newSign$ **then**
            return p_i
        end if
    end for
    $offset \leftarrow offset + \text{Length}[\textbf{P}]$
end for

We will now present the results obtained by implementing and running our algorithm in Wolfram Mathematica 11.1.

III. RESULTS

Using the methods described in Section II. we found Skewes numbers for 8 new prime k-tuples. Tab. 2 shows these numbers.

Various other results emerge from the above computations. First, we find that the inequality $\pi_P(n) < C_{a_1,a_2,...,a_k} \int_2^n \frac{dt}{\log^{k+1} t}$ only holds within a short interval for all of the prime tuples that have been considered. Taking for instance P_{4a}, it appears that the reversal of this in-

equality remains true after merely the 9th sign change of

$$\pi_{P_{4a}}(n) - C_{P_{4a}} \int_2^n \frac{dt}{\log^4 t}.$$

Tab. 2. The Skewes numbers for the prime k-tuples considered in our study

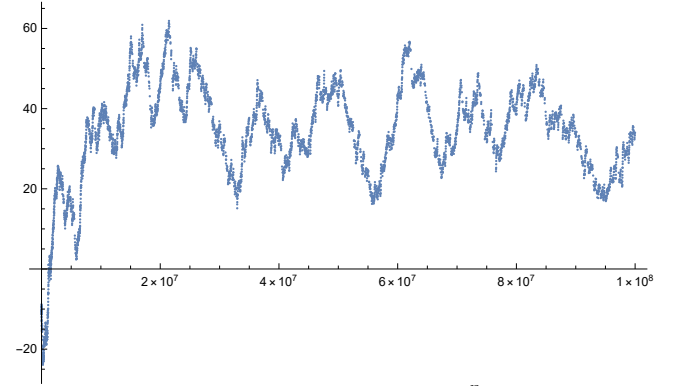| Prime k-tuple	Skewes number
P_{2a}	1369391
P_{2b}	5206837
P_{3a}	87613571
P_{3b}	337867
P_{4a}	1172531
P_{4b}	827929093
P_{5a}	21432401
P_{5b}	216646267
P_6	251331775687

Fig. 1. Value of $\pi_{P_{4a}}(n) - C_{P_{4a}} \int_2^n \frac{dt}{\log^4 t}$ within the interval $[2, 10^8]$
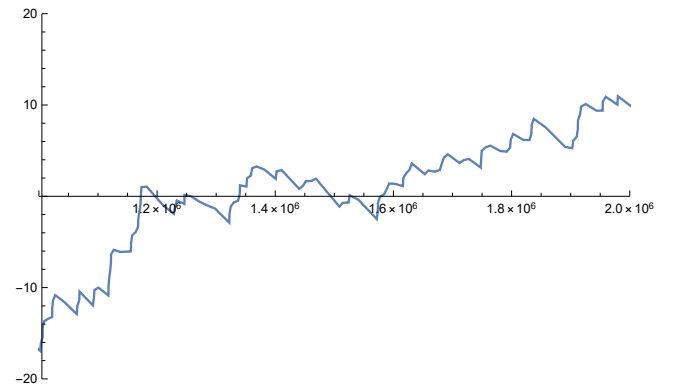

Fig. 2. The crossover region in P_{4a}

Tab. 3. Values of $\pi_{P_6}(n)$ within intervals of size 10^{10} and the corresponding Hardy-Littlewood estimate (computed to two significant figures after the decimal point)

Interval ($[i, j] \times 10^{10}$)	$\pi'_{P_6}(i, j) = \pi_{P_6}(j) + -\pi_{P_6}(i)$	Hardy-Littlewood estimate $\epsilon(i, j)$	$\epsilon(i, j)/\pi'_{P_6}(i, j)$
$[0, 1]$	1613	1664.55	1.03195
$[1, 2]$	1065	1052.01	0.98780
$[2, 3]$	897	920.42	1.02610
$[3, 4]$	813	845.55	1.04003
$[4, 5]$	796	794.54	0.99816
$[5, 6]$	759	756.47	0.99666
$[6, 7]$	674	726.43	1.07778
$[7, 8]$	680	701.82	1.03208
$[8, 9]$	680	681.10	1.00161
$[9, 10]$	649	663.29	1.02201
$[10, 11]$	638	647.73	1.01525
$[11, 12]$	628	633.96	1.00949
$[12, 13]$	610	621.64	1.01908
$[13, 14]$	608	610.52	1.00414
$[14, 15]$	605	600.40	0.99239
$[15, 16]$	612	591.13	0.96589
$[16, 17]$	613	582.59	0.95039
$[17, 18]$	601	574.69	0.95622
$[18, 19]$	620	567.34	0.91506

Fig. 1 shows a plot of this difference within the interval $[2, 10^8]$ while Fig. 2 zooms in on the crossover region. In another example, we observed a similar behaviour for P_6, which showed 15 sign changes between its Skewes number (251331775687) and 26×10^{10} (the Skewes number included).

Our computations yielded other intriguing results as well. Indeed, it follows from the first Hardy-Littlewood conjecture that cousin primes ($P_{2b} = (p, p + 4)$) have the same asymptotic density as twin primes; however, the Skewes number for cousin primes is almost 4 times greater than its twin prime counterpart. This is similar to the tuples P_{5a} and P_{5b} which, despite their equal conjectured density, have their Skewes number far apart; indeed, the one for P_{5b} is almost 10 times larger than that for P_{5a}.

Another consequence of the first Hardy-Littlewood conjecture is the thinning out of prime tuples within intervals of equal size. The results in our study do indeed provide evidence for such behaviour. Here we give the example of P_6 by comparing the value of $\pi_{P_6}(n)$ with the Hardy-Littlewood estimate (computed here to two significant figures after the

decimal point) within intervals of size 10^{10}. Tab. 3 shows our numerical results.

Finally, based on the results in our study, we propose the following conjecture.

Conjecture 3 *All admissible prime k-tuples have a Skewes number.*

IV. CONCLUSION AND FURTHER WORK

Alongside the discovery of 8 new Skewes numbers for prime k-tuples, our results also give further evidence for the validity of the first Hardy-Littlewood conjecture. However, we find that tuples that are expected to have the same asymptotic density do not behave in the same manner, for instance their Skewes numbers occur considerable distances apart.

Since our numerical data does not reach a sufficiently high upper bound, we are unable to investigate the validity of Wolf's conjecture within the context of other prime tuples. For instance, the cousin primes are expected to have the same density as the twin primes, thus according to Wolf's conjecture the amount of sign changes in the difference $\pi_{P_{2b}}(n) - C_{P_{2b}}\mathrm{Li}_2(n)$ within the interval $n \in (1, T)$ should also be asymptotic to $\dfrac{\sqrt{T}}{\log T}$.

We would also find interesting the extension of the search for Skewes numbers to further k-tuples such as 2-tuples of the type $(p, p + 2k)$ for positive integer $k > 2$ (the case $k = 3$ often being referred to as the "sexy primes" within more relaxed contexts), or longer ones such as

- $P_{7a} = (p, p + 2, p + 6, p + 8, p + 12, p + 18, p + 20)$ and
- $P_{7b} = (p, p + 2, p + 8, p + 12, p + 14, p + 18, p + 20)$.

Both of these tuples are expected to have the same asymptotic density,

$$\pi_{7a}(n), \pi_{7b}(n) \sim \frac{35^6}{3 \times 2^{22}} \prod_{p \geq 11} \frac{p^6(p - 7)}{(p - 1)^7} \int_2^n \frac{dt}{\log^7 t},$$

but we did not find any primes that violate the Hardy-Littlewood inequality up to 1.2×10^{11}.

References

[1] R.P. Brent, *Irregularities in the distribution of primes and twin primes*, Math. Comp. **29**, 43–56 (1975).

[2] M. Wolf, *The Skewes number for twin primes: counting sign changes of $\pi_2(x) - C_2\mathrm{Li}_2(x)$*, Comput. Methods Sci. Technol. **17**, 87–92 (2011).

[3] M. Wolf, *Random walk on the prime numbers*, Physica A **250**, 335–344 (1998).

[4] Th.R. Nicely, *New evidence for the infinitude of some prime constellations*, (2004). http://www.trnicely.net/ipc/ipc1d.pdf

[5] G.H. Hardy, J.E. Littlewood, *Some problems of 'Partitio Numerorum' III: On the expression of a number as a sum of primes*, Acta Math. **44**, 1–70 (1922).

[6] J.E. Littlewood, *Sur la distribution des nombres premiers*, C. R. Math. Acad. Sci. Paris **158**, 1869–1872 (1914).

[7] S. Skewes, *On the difference $\pi(x) - \mathrm{li}(x)$*, J. London Math. Soc. **8**, 277–283 (1933).

[8] S. Skewes, *On the difference $\pi(x) - \mathrm{li}(x)$ (II)*, Proc. London Math. Soc. **5**, 48–70 (1955).

László Tóth holds a Master's degree in Computer Science from the University of East Anglia. He is passionate about mathematics, in particular the fields of analytic and computational number theory, in which he has already published. He also holds strong interest in automatic sequences (such as the Thue-Morse sequence) and related analytic functions. He works in the Grand Duchy of Luxembourg since 2013 as a software developer.