NEW BOUNDS AND COMPUTATIONS ON PRIME-INDEXED PRIMES

Jonathan Bayless  
Department of Mathematics, Husson University, Bangor, Maine  
baylessj@husson.edu

Dominic Klyve  
Dept. of Mathematics, Central Washington University, Ellensburg, Washington  
klyved@cwu.edu

Tomás Oliveira e Silva  
Electronics, Telecommunications, and Informatics Department, Universidade de Aveiro, Aveiro, Portugal  
tos@ua.pt

Received: 5/25/12, Revised: 5/15/13, Accepted: 5/16/13, Published: 7/10/13

Abstract
In a 2009 article, Barnett and Broughan considered the set of prime-index primes. If the prime numbers are listed in increasing order (2, 3, 5, 7, 11, 13, 17, ...), then the prime-index primes are those which occur in a prime-numbered position in the list (3, 5, 11, 17, ...). Barnett and Broughan established a prime-indexed prime number theorem analogous to the standard prime number theorem and gave an asymptotic for the size of the $n$-th prime-indexed prime.

We give explicit upper and lower bounds for $\pi^2(x)$, the number of prime-indexed primes up to $x$, as well as upper and lower bounds on the $n$-th prime-indexed prime, all improvements on the bounds from 2009. We also prove analogous results for higher iterates of the sequence of primes. We present empirical results on large gaps between prime-index primes, the sum of inverses of the prime-index primes, and an analog of Goldbach’s conjecture for prime-index primes.

1. Introduction
Many of the classes of primes typically studied by number theorists concern properties of primes dictated by the positive integers. Twin primes, for example, are consecutive primes with an absolute difference of 2. It is interesting, however, to consider a sequence of primes whose members are determined by the primes themselves.
In this work, we consider the set of prime-indexed-primes (or PIPs), which are prime numbers whose index in the increasing list of all primes is itself prime. In particular we have:

**Definition 1.1.** Let \( P \) be the sequence of primes written in increasing order \( \{p_i\}_{i \geq 1} \). The sequence of prime-indexed-primes, or PIPs, is the subsequence of \( P \) where the index \( i \) is itself prime. Specifically, the sequence of prime-indexed primes is given by \( \{q_i\} \) where \( q_i = p_{p_i} \) for all \( i \geq 1 \).

Prime-indexed-primes seem to have been first considered in 1965 by Jordan [1], who also considered primes indexed by other arithmetic sequences. They were again studied in 1975 by Dressler and Parker [2], who showed that every positive integer greater than 96 is representable by the sum of distinct PIPs. Later, Sándor [3] built on Jordan’s work of considering the general question of primes indexed by an arithmetic sequence by studying reciprocal sums of such primes and limit points of the difference of two consecutive primes. Some of these results can be found in [4, p. 248-249].

The direct inspiration of this work is the recent work of Broughan and Barnett [5], who have demonstrated many properties of PIPs, giving bounds on the \( n \)-th PIP, a PIP counting function (analogous to the prime number theorem), and some results on small gaps between consecutive PIPs.

We follow Broughan and Barnett by explaining PIPs via example. In the following list of primes up to 109, all those with prime index (the PIPs) are in bold type: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109.

We note also the equivalent definition of PIPs, which may be of some use. If we let, as usual, \( \pi(x) \) be the number of primes not greater than \( x \), then an integer \( p \) is a PIP if both \( p \) and \( \pi(p) \) are prime.

2. Preliminary Lemmas

In this section we state a few lemmas which will be helpful in the proofs of the theorems in the remaining sections of this paper. The proofs consist of technical details, and contain no insight of direct relevance to the rest of the paper, so we defer their proofs to Section 10. First, we give a pair of bounds comparing combinations of \( n \) and \( \log n \) to \( \log \log n \).

**Lemma 2.1.** For \( n \geq 3 \),

\[
\log (\log (n \log n)) < \log \log n + \frac{\log \log n}{\log n}.
\]
Lemma 2.2. For $n \geq 3$,
\[
\log \log (n \log (n \log n)) < \left(1 + \frac{1}{\log n} + \frac{1}{\log^2 n}\right) \log \log n.
\]

We will also need two lemmas bounding the reciprocal of $\log \pi(x)$, one for explicit bounds and one for asymptotic bounds.

Lemma 2.3. For all $x \geq 33$,
\[
\frac{1}{\log x} \left(1 + \frac{\log \log x}{\log x}\right) < \frac{1}{\log \pi(x)} < \frac{1}{\log x} \left(1 + \log \frac{\log x}{\log^2 x} + \frac{(\log \log x)^2}{\log^2 x} + \cdots\right).
\]

Lemma 2.4. For all $k \geq 1$, we have
\[
\frac{1}{\log^k \pi(x)} = \frac{1}{\log^k x} \left(1 + \frac{k \log \log x}{\log x} + O_k \left(\frac{(\log \log x)^2}{\log^2 x}\right)\right)
\]
and
\[
\frac{\log \log \pi(x)}{\log \pi(x)} = \frac{\log \log x}{\log x} + O \left(\frac{(\log \log x)^2}{\log^2 x}\right).
\]

3. Bounds on PIPs

Broughan and Barnett were the first to put upper and lower bounds on $q_n$, the $n$-th PIP. In particular, they show
\[
q_n < n (\log n + 2 \log \log n)(\log n + \log \log n) - n \log n + O(n \log \log n),
\]
\[
q_n > n (\log n + 2 \log \log n)(\log n + \log \log n) - 3n \log n + O(n \log \log n).
\]

By using some theorems of Dusart [6], we show that these bounds can both be improved and made explicit.

Theorem 3.1. For all $n \geq 3$, we have
\[
q_n > n (\log n + \log \log n - 1)(\log n + 2 \log \log n - 1),
\]
and for all $n \geq 71$, we have
\[
q_n < n \left(\log n + \log \log n - \frac{1}{2}\right) \left(\log n + 2 \log \log n + \frac{3 \log \log n}{\log n} - \frac{1}{2}\right).
\]

Proof. We begin with the lower bound. Using the result of Dusart [6] that
\[
p_n > n (\log n + \log \log n - 1)
\]
for all \( n \geq 2 \), we have for all \( n \geq 3 \) that
\[
q_n > n (\log n + \log \log n - 1) \left( \log (n (\log n + \log \log n - 1))
+ \log \log (n (\log n + \log \log n - 1)) - 1 \right).
\]

Now, using the fact that
\[
\log (n \log n + n \log \log n - n) > \log n + \log \log n,
\]
we may bound
\[
q_n > n (\log n + \log \log n - 1) (\log n + \log \log n + \log (\log n + \log \log n) - 1)
> n (\log n + \log \log n - 1) (\log n + 2 \log \log n - 1).
\]

Turning to the upper bound, we use a result of Rosser and Schoenfeld [7], that:
\[
p_n < n \left( \log n + \log \log n - \frac{1}{2} \right)
\]
for \( n \geq 20 \). To simplify the notation here, let
\[
N = n \left( \log n + \log \log n - \frac{1}{2} \right).
\]

This gives that for \( n \geq 71 \),
\[
q_n < N \left( \log N + \log \log N - \frac{1}{2} \right).
\]

Now, Lemma 2.1 gives that
\[
\log N < \log n + \log \log n + \frac{\log \log n}{\log n}.
\]

Also, by Lemma 2.2,
\[
\log \log N < \left(1 + \frac{1}{\log n} + \frac{1}{\log^2 n}\right) \log \log n < \log \log n + \frac{2 \log \log n}{\log n}.
\]

Putting these into (3.2), we can bound
\[
q_n < n \left( \log n + \log \log n - \frac{1}{2} \right) \left( \log n + 2 \log \log n + \frac{3 \log \log n}{\log n} - \frac{1}{2} \right),
\]
which is the bound in the theorem. \( \square \)
In fact, using an upper bound on \( p_n \) due to Robin [8], namely
\[
p_n < n \left( \log n + \log \log n - 0.9385 \right) \text{ for } n \geq 7022,
\]
we are able to prove the somewhat stronger upper bound
\[
q_n < n \left( \log n + \log \log n - 0.9385 \right) \left( \log n + 2 \log \log n + \frac{2 \log \log n}{\log n} - 0.9191 \right)
\]
for all \( n \geq 70919 \).

4. Bounds for the Number of Prime-Indexed-Primes

Define \( \pi^2(x) \) to be the number of PIPs not greater than \( x \). Note that it follows immediately from the definition that \( \pi^2(x) = \pi(\pi(x)) \). Broughan and Barnett show that
\[
\pi^2(x) \sim \frac{x}{\log^2 x},
\]
and also give the slightly more sophisticated
\[
\pi^2(x) \sim \frac{x}{\log^2 x} + O \left( \frac{x \log \log x}{\log^3 x} \right).
\]

With respect to the classical prime number theorem, a study of the error term has driven much research in number theory. Among the known results [6, page 16] is
\[
\pi(x) = \frac{x}{\log x} \left( 1 + \frac{1}{\log x} + \frac{2}{\log^2 x} + O \left( \frac{1}{\log^4 x} \right) \right).
\]

We also have a number of known explicit bounds on \( \pi(x) \), including [6, page 2]
\[
\frac{x}{\log x} \left( 1 + \frac{1}{\log x} \right) \leq \pi(x) \leq \frac{x}{\log x} \left( 1 + \frac{1.2762}{\log x} \right),
\]
where the lower bound holds for \( x \geq 599 \) and the upper bound holds for \( x > 1 \). Based on this work, we are able to prove the following theorem.

**Theorem 4.1.** For all \( x \geq 3 \), we have
\[
\pi^2(x) < \frac{x}{\log^2 x} \left( 1 + \frac{1.5}{\log x} \right)^2 \left( 1 + \frac{\log \log x}{\log x} + \frac{1.5(\log \log x)^2}{\log^2 x} \right),
\]
while, for all \( x \geq 179 \),
\[
\pi^2(x) > \frac{x}{\log^2 x} \left( 1 + \frac{1}{\log x} \right)^2 \left( 1 + \frac{\log \log x}{\log x} \right).
\]
Proof. We begin with the upper bound. First, note that
\[
\sum_{i=2}^{\infty} \left( \frac{\log \log x}{\log x} \right)^i = \left( \frac{\log \log x}{\log x} \right)^2 \left( 1 - \frac{\log \log x}{\log x} \right) < 1.5 \left( \frac{\log \log x}{\log x} \right)^2
\]  
(4.3)
for \( x \geq 179 \). We may bound \( \pi^2(x) \) with bounds on \( \pi(x) \), using Lemma 2.3 and (4.3) to see that
\[
\pi^2(x) \leq \frac{\pi(x)}{\log \pi(x)} \left( 1 + \frac{1.2762}{\log \pi(x)} \right)
\]
\[
< \frac{x}{\log^2 x} \left( 1 + \frac{1.2762}{\log x} \right) \left( 1 + \frac{1.2762}{\log \pi(x)} \right) \left( 1 + \frac{\log \log x}{\log x} + \frac{1.5(\log \log x)^2}{\log^2 x} \right).
\]
Now, to complete the upper bound’s proof, we need only show
\[
\left( 1 + \frac{1.2762}{\log x} \right) \left( 1 + \frac{1.2762}{\log \pi(x)} \right) < \left( 1 + \frac{1.5}{\log x} \right)^2.
\]  
(4.4)
By Lemma 2.3 and (4.3),
\[
\frac{1.2762}{\log \pi(x)} \leq \frac{1.2762}{\log x} \left( 1 + \frac{\log \log x}{\log x} + \frac{1.5(\log \log x)^2}{\log^2 x} \right) < \frac{1.7238}{\log x},
\]
for \( x \geq 3030 \). Then,
\[
\left( 1 + \frac{1.2762}{\log x} \right) \left( 1 + \frac{1.7238}{\log x} \right) \leq 1 + \frac{3}{\log x} + \frac{2.22}{\log^2 x} < \left( 1 + \frac{1.5}{\log x} \right)^2,
\]
which establishes (4.4) and thus the upper bound in the theorem for \( x \geq 3030 \). Finally, a computer check verifies the upper bound for \( 3 \leq x \leq 3030 \).

For the lower bound, we can use (4.2) to see that
\[
\pi^2(x) \geq \frac{\pi(x)}{\log \pi(x)} \left( 1 + \frac{1}{\log \pi(x)} \right)
\]
\[
\geq \frac{\pi(x)}{\log \pi(x)} \left( 1 + \frac{1}{\log x} \right)
\]
\[
\geq \frac{x}{\log \pi(x)} \left( 1 + \frac{1}{\log x} \right)^2 \cdot \frac{1}{\log \pi(x)}
\]
\[
\geq \frac{x}{\log^2 x} \left( 1 + \frac{1}{\log x} \right)^2 \left( 1 + \frac{\log \log x}{\log x} \right)
\]
by Lemma 2.3, for all \( x \geq 4397 \). A computer check shows that the bound also holds for all \( 179 \leq x \leq 4397 \), completing the proof of the theorem. \( \square \)
Platt
Non
computed
xagariasTyillerT–dlyzk
all
and
also
Theorem
is
to
of
INTEGERS:

Small improvements can be made to the bounds in Theorem 4.1, at the expense of the relative simplicity in the statement. However, the lower bound is much closer to the truth. Using a stronger version of (4.2) and a finer version of Lemma 2.3 it is possible to prove the following theorem.

**Theorem 4.2.** We have

\[ \pi^2(x) = \frac{x}{\log^2 x} \left(1 + \frac{\log \log x}{\log x} + \frac{2}{\log x}\right) + O\left(\frac{x(\log \log x)^2}{\log^4 x}\right). \]

Table 4 gives the values of \( \pi(x) \) and \( \pi^2(x) \) for powers of 10 up to \( 10^{24} \). The table also gives values of \( \pi^3(x) \), the number of PIPs of prime index up to \( x \) (see a definition and further generalization in Section 7). The values of \( \pi(x) \) up to \( 10^{23} \), as well as all values of \( \pi^2(x) \) and of \( \pi^3(x) \), were computed using an implementation of the Lagarias-Miller-Odlyzko algorithm [9] described in [10]. The value of \( \pi(10^{24}) \) was computed in 2010 by Buethé, Franke, Jost, and Kleinjung [11] using a conditional (on the Riemann hypothesis) analytic method, and latter confirmed in 2012 by Platt [12] using an unconditional analytic method.
5. Twin PIPs

Twin primes are pairs of primes which are as close together as possible – namely, they are consecutive odd numbers which are both prime. We will define twin prime-indexed primes, or twin PIPs, with a similar motivation, that is, they are PIPs which are as close together as possible. Since two consecutive indices cannot be prime, there must always be at least one prime between consecutive PIPs (except for the trivial case \( q_2 = 3, q_3 = 5 \)). Furthermore, we know that \( p, p + 2, \) and \( p + 4 \) cannot be simultaneously prime, since 3 always divides one of them, and thus the smallest distance between consecutive PIPs is 6. To this end, we make the following definition.

**Definition 5.1.** Let \( p \) and \( q \) be PIPs. We say that they are twin PIPs if \( p - q = 6 \).

The twin prime conjecture states that there are infinitely many twin primes, and Broughan and Barnett conjecture that there are infinitely many consecutive PIPs with a difference of 6 (and indeed they conjecture that all gaps of even size at least 6 appear infinitely often). However, the Twin Prime Conjecture has a strong form, which states that where \( \pi_2(x) \) is the number of twin primes not greater than \( x \),

\[
\pi_2(x) \sim 2C_{\text{twin}} \int_0^x \frac{dt}{\log^2 t} \sim 2C_{\text{twin}} \frac{x}{\log^2 x},
\]

where

\[
C_{\text{twin}} = \prod_{p \geq 3} \frac{p(p - 2)}{(p - 1)^2} \approx 0.6601618158
\]

is the twin prime constant. One of the reasons that number theorists have faith in the twin prime conjecture is that the strong form of the conjecture seems to be very accurate. Experimentally we see that the data conform to this conjecture remarkably closely. The twin prime conjecture predicts that the number of primes up to \( 4 \times 10^{18} \) is about

\[
2C_{\text{twin}} \int_0^{4 \times 10^{18}} \frac{dx}{\log^2 x} \approx 3023463139207178.4.
\]

The third author has calculated this value precisely [13], and found

\[
\pi_2(4 \times 10^{18}) = 3023463123235320,
\]

which impressively agree in the first eight digits.

The basic idea behind the strong form of the twin prime conjecture is that the events “\( p \) is prime” and “\( p + 2 \) is prime” are not independent events (details can be found in, say, [14, pp. 14–16]). Similar reasoning applies to twin PIPs. In particular, we consider the following question: if \( q \) is a PIP, what is the probability
that \( q + 6 \) is also a PIP? We must have either that \( q, q + 2, q + 6 \) are all prime, or that \( q, q + 4, q + 6 \) are all prime. But it is not enough for \( q + 6 \) to be prime; it must also be a PIP. Combining these ideas, we find the following

**Theorem 5.2.** If a prime \( q \) with index \( n \) is the first of a pair of twin PIPs, then one of two cases hold. Either: The triple \((q, q+2, q+6)\) are all prime, or the triple \((q, q+4, q+6)\) are all prime. Furthermore, the index of \( q \) and \( q + 6 \) must each be prime.

From this theorem we can construct a heuristic bound on the density of twin PIPs.

**Conjecture 5.3.** The number of twin PIPs up to \( x \), \( \pi_2^2(x) \), is asymptotically

\[
\prod_{p>3} \frac{p^3(p-2)(p-3)}{(p-1)^5} \cdot \int_2^x \frac{dt}{\log^3 t (\log t - \log \log t)^2}.
\]

**Argument.** Hardy and Little established conjectures on the density of prime constellations a century ago [15], and though unproven they are widely accepted, and enjoy considerable empirical support. Their conjectured density of either triple given in Theorem 5.2 is asymptotically

\[
P_x(q, q+2, q+6) \sim \prod_{p>3} \frac{p^2(p-3)}{(p-1)^3} \cdot \int_2^x \frac{dt}{\log^3 t}.
\]

We expect, as we have throughout this paper, that we can treat the primality of the index of a prime \( q \) as independent of \( q \). Let \( n \) be the index of \( q \). Then \( n + 2 \) is the index of the prime \( q + 6 \). The probability of both \( n \) and \( n + 2 \) being prime is heuristically

\[
\prod_{p>2} \frac{p(p-2)}{(p-1)^2} \cdot \frac{1}{\log^2 n} = \prod_{p>2} \frac{p(p-2)}{(p-1)^2} \cdot \frac{1}{\log^2 (q/\log q)}.
\]

If we simply multiply our earlier heuristic by the probability of this additional restriction, we find an expected density of twin PIPs to be

\[
L(x) = \prod_{p>3} \left( \frac{p^2(p-3)}{(p-1)^3} \right) \left( \frac{p(p-2)}{(p-1)^2} \right) \int_2^x \frac{dt}{\log^3 t \log^2 (t/\log t)}
\]

\[
= \prod_{p>3} \frac{p^3(p-2)(p-3)}{(p-1)^5} \int_2^x \frac{dt}{\log^3 t (\log t - \log \log t)^2}.
\]

Evaluating the product gives

\[
L(x) \approx 7.5476417 \int_2^x \frac{dt}{\log^3 t (\log t - \log \log t)^2}.
\]

This heuristic seems to describe the distribution of twin PIPs quite well. The following table gives the predicted number and actual number of twin PIPs up to various powers of 10, together with the absolute and relative error at each stage.
6. The Sum of the Reciprocals of the PIPs

The asymptotic density of the PIPs is $O(x/\log^2 x)$, from which it follows that the sum of the reciprocals of the PIPs converges (as noted first in [1]). Reciprocal sums have some interest in themselves; bounding Brun’s constant, the sum of the reciprocals of the twin primes, has been a goal of many mathematicians since at least 1974 [16, 17, 18, 19, 20]. However, the accuracy of bounds on reciprocal sums also measures in an important way how much we understand a particular class of numbers. Bounding a reciprocal sum well requires two things: first, a computationally determined bound on small integers from the class; and second, good explicit bounds on the density of large integers from the class.

By this measure, twin primes are understood much better than, say, amicable numbers. Let $B$ be the sum of the reciprocals of the twin primes. Then $B$ has been shown to satisfy [14]

$$1.83 < B < 2.347,$$

and in fact [20]

$$1.83 < B < 2.15,$$

assuming the Extended Riemann Hypothesis. By contrast, let $P$ be the Pomerance Constant – the sum of the reciprocals of the amicable numbers. Then the best known bounds on $P$ [21] are the fairly weak

$$0.01198 < P < 6.56 \times 10^8.$$
Comparing the size of these intervals shows that, in a measurable way, current mathematical knowledge about twin primes is better than that of amicable numbers. With this in mind, we are interested in determining the accuracy to which we can bound the sum of the reciprocals of the PIPs.

By generating all PIPs up to $10^{15}$ using the obvious expedient of employing two segmented Eratosthenes sieves [22], one to check the primality of each odd $n$ in that interval and another to check the primality of $\pi(n)$, and by accumulating the sum of the inverses of the PIPs using a 192-bits fractional part, it we find that

$$\sum_{q \leq 10^{15}} \frac{1}{q} \approx 1.01243131879802898253.$$  \hspace{1cm} (6.1)

When $u$ and $v$ are not PIPs, the sum

$$T(u, v) = \sum_{u < q < v} \frac{1}{q} = \int_u^v \frac{d\pi(\pi(x))}{x}$$

can be reasonably well approximated by replacing $\pi(x)$ by $\text{li}(x) = \int_0^x \frac{dt}{\log t}$. This yields

$$T(u, v) \approx \hat{T}(u, v) = \int_u^v \frac{dx}{x \log x \log \text{li}(x)} = \int_{\log u}^{\log v} \frac{dx}{x \log \text{li}(e^x)}.$$ 

Due to potential arithmetic overflow problems when $x$ is large, in this last integral $\log \text{li}(e^x)$ should be evaluated by replacing it by its asymptotic expansion

$$\text{li}(e^x) \approx x - \log x + \log \left( \sum_{k=0}^N \frac{k!}{x^k} \right);$$

$N = \lfloor x \rfloor$ delivers an approximation with relative error close to $\sqrt{2\pi x e^{-x}}$. This is not enough to evaluate $T(10^{15}, \infty)$ directly with an absolute error smaller that $5 \times 10^{-21}$, so $\hat{T}(10^{15}, 10^{100})$ can be numerically integrated without using the asymptotic expansion, and then $T(10^{100}, \infty)$ can be numerically integrated using the asymptotic expansion. Both Mathematica and pari-gp agree that

$$\hat{T}(10^{15}, \infty) = 0.03077020549198786752.$$ 

It follows that

$$\sum_{q} \frac{1}{q} \approx 1.04320152429001685005.$$  \hspace{1cm} (6.2)

Comparing $T(k \times 10^{14}, (k+1) \times 10^{14})$ with $\hat{T}(k \times 10^{14}, (k+1) \times 10^{14})$ for $k = 1, \ldots, 10$ suggests that the absolute value of the relative error of the latter is, with high probability, smaller in $10^{-6}$. The error of our estimate of $\sum_{q} \frac{1}{q}$ is therefore expected to be of order $10^{-8}$.
Another approach to finding the sum of the reciprocals of the PIPs is to use the explicit upper and lower bounds on \( \pi^2(x) \) from Theorem 4.1, our calculations to \( 10^{15} \), and partial summation to bound the sum. Let us label the upper bound on \( \pi^2(x) \) as \( \pi_{up}^2(x) \), and the lower bound as \( \pi_{lb}^2(x) \). Then we have

\[
\sum_{q \leq 10^{15}} \frac{1}{q} \sum_{q \leq 10^{15}} \frac{1}{q} \left( \frac{\pi^2(10^{15})}{10^{15}} \right) + \int_{10^{15}}^{\infty} \frac{\pi_{up}^2(t)}{t^2} dt,
\]

and

\[
\sum_{q \leq 10^{15}} \frac{1}{q} \sum_{q \leq 10^{15}} \frac{1}{q} \left( \frac{\pi^2(10^{15})}{10^{15}} \right) + \int_{10^{15}}^{\infty} \frac{\pi_{lb}^2(t)}{t^2} dt.
\]

In fact, these functions \( \pi_{up}^2(x) \) and \( \pi_{lb}^2(x) \) do a fairly good job bounding \( \pi^2(x) \) past \( 10^{15} \), as determined by the difference in the integrals above. Numerical calculation with Mathematica gives the following:

\[
\int_{10^{15}}^{\infty} \frac{\pi_{up}^2(t)}{t^2} dt \approx 0.0315569; \quad \int_{10^{15}}^{\infty} \frac{\pi_{lb}^2(t)}{t^2} dt \approx 0.0322135.
\]

Combining these values with the calculation in (6.1), we can show that the sum of the reciprocals of the PIPs satisfies

\[1.04299 < \sum_{q} \frac{1}{q} < 1.04365,\]

in good agreement with (6.2).

7. The Generalized Prime Number Theorem

It is natural to consider a further generalization of prime-index primes. If the set of primes is listed in order, the subsequence of prime-index primes could be called 2-primes. Similarly, if the set of 2-primes is listed in increasing order, we may call the subsequence with prime index 3-primes. Let a \( k \)-prime be a member of the \( k \)-th iteration of this process. One may ask for the analogous results on the \( n \)-th \( k \)-prime and the number of \( k \)-primes up to \( x \).

As noted in Broughan and Barnett [5], it is not hard to establish an analog to the Prime Number Theorem. Namely, defining \( \pi^k(x) \) as the number of \( k \)-primes less than or equal to \( x \), it is easy to show that

\[\pi^k(x) = \frac{x}{\log^k x} + O\left(\frac{x \log \log x}{\log^k x}\right).\]

In fact, it can be shown that \( \pi^k(x) \sim \text{Li}^k(x) \) as \( x \to \infty \). The proof of this statement is not difficult, but is also not very enlightening. The theorem we prove here
is slightly weaker, but shows more of the shape of the three main terms in this asymptotic.

**Theorem 7.1.** For all $k \geq 1$,

$$
\pi^k(x) = \frac{x}{\log^k x} \left(1 + \frac{(k-1) \log \log x}{\log x} + \frac{k}{\log x} \right) + O_k \left( \frac{x(\log \log x)^2}{\log^{k+2} x} \right).
$$

**Proof.** The proof proceeds by induction on $k$. The case $k = 1$ is given in (4.1), so we assume the statement holds up to $k$. Then,

$$
\pi^{k+1}(x) = \pi^k(\pi(x))
$$

$$
= \frac{\pi(x)}{\log^k \pi(x)} \left(1 + \frac{(k-1) \log \log \pi(x)}{\log \pi(x)} + \frac{k}{\log \pi(x)} \right)
$$

$$
+ O_k \left( \frac{\pi(x)(\log \log \pi(x))^2}{\log^{k+2} \pi(x)} \right)
$$

by the induction hypothesis. Now, Lemma 2.4 gives that

$$
\frac{1}{\log^k \pi(x)} \left(1 + \frac{(k-1) \log \log \pi(x)}{\log \pi(x)} + \frac{k}{\log \pi(x)} \right)
$$

is equivalent to

$$
\frac{1}{\log^k x} \left(1 + \frac{k \log \log x}{\log x} + \frac{k}{\log x} + O_k \left( \frac{(\log \log x)^2}{\log^2 x} \right) \right).
$$

Putting this together with (4.1) gives that

$$
\pi^{k+1}(x) = \frac{\pi(x)}{\log^k \pi(x)} \left(1 + \frac{k \log \log \pi(x)}{\log \pi(x)} + \frac{k}{\log \pi(x)} + O_k \left( \frac{(\log \log \pi(x))^2}{\log^2 \pi(x)} \right) \right)
$$

$$
= \frac{x}{\log^{k+1} x} \left(1 + \frac{k \log \log x}{\log x} + \frac{k + 1}{\log x} \right) + O_k \left( \frac{x(\log \log x)^2}{\log^{k+2} x} \right),
$$

completing the theorem’s proof. 

Using Theorem 4.1 as the base case and Lemma 2.3 for the inductive step, we can prove the following theorem giving explicit bounds on $\pi^k(x)$.

**Theorem 7.2.** For all $k \geq 2$, there exists a computable $x_0(k)$ such that

$$
\pi^k(x) < \frac{x}{\log^k x} \left(1 + \frac{1.5 \log \log x}{\log x} \right)^k \left(1 + \frac{1.5 \log \log x}{\log x} \right)^{k-1}
$$

and

$$
\pi^k(x) > \frac{x}{\log^k x} \left(1 + \frac{1 \log \log x}{\log x} \right)^k \left(1 + \frac{\log \log x}{\log x} \right)^{k-1}
$$

for all $x \geq x_0(k)$. 
Note that from the proof of Lemma 2.3 and Theorem 4.1 we may choose any $x_0(k)$ satisfying $\pi^k(x_0(k)) \geq 13$, as $\pi^2(179) = 13$.

It is also not difficult to adapt an argument from [5] to prove a proposition on $\pi^k(x)$ for all $k \geq 1$.

**Proposition 7.3.** The following inequalities are true for every integer $n > 1$ and $k \geq 1$ and for all sufficiently large real numbers $x, y$:

(a) $\pi^k(nx) < n\pi^k(x)$,
(b) $\pi^k(x + y) \leq \pi^k(x) + 2^k\pi^k(y)$, and
(c) $\pi^k(x + y) - \pi^k(x) \leq k \frac{\log y}{\log x}$.

**Proof.** Each of the statements in this theorem have been shown for $k = 2$ in [5] and for $k = 1$ elsewhere ((a) in [23], (b) and (c) in [24]). We assume these base cases and follow the argument in [5] to complete the induction for each statement.

(a) From Panaitopol [23], we have $\pi(nx) < n\pi(x)$ for sufficiently large $x$. The induction hypothesis, followed by an application of Panaitopol’s result gives

$$\pi^{k+1}(nx) < \pi(n\pi^k(x)) < n\pi^{k+1}(x)$$

for sufficiently large $x$.

(b) Using Montgomery and Vaughan’s [24] bound $\pi(x+y) \leq \pi(x) + 2\pi(y)$ together with the induction hypothesis, we have

$$\pi^{k+1}(x + y) \leq \pi(\pi^k(x) + 2^k\pi(y)) \leq \pi^{k+1}(x) + 2^{k+1}\pi(y)$$

for sufficiently large $x$ and $y$.

(c) This follows from part (b) and Theorem 7.1. \qed

Note that these bounds are certainly not the best possible. Inequality (b) in particular seems rather weak. Proving a stronger general theorem, however, seems difficult.

8. Gaps Between PIPs

Our consideration in Section 5 of twin PIPs, or consecutive PIPs with difference 6, is just a special case of a more general question about gaps between consecutive PIPs. In this section we consider some computational data on other gap sizes. Let

$$q(h) = \min_{q_{i+1} - q_i = h} q_i$$
be the first occurrence of a gap of size $h$ between PIPs (or infinity if no such gap exists), let

$$Q(x; h) = \sum_{q_i \leq x \atop q_{i+1} - q_i = h} 1$$

be the number of gaps of $h$ between PIPs up to $x$, and let

$$F(h) = \prod_{p > 2 \atop p \mid h} \frac{p - 1}{p - 2}$$

be the corresponding Hardy-Littlewood correction factor. As expected due to the prime $k$-tuples conjecture, it was found that the graph of $Q(10^{15}; h)$ exhibited a rapid "oscillation" (cf., for example, [25]), which disappeared in a graph of $Q(10^{15}; h)/F(h)$. Contrary to what happens with the graphs of smoothed counts of prime gaps (i.e., counts divided by $F(h)$), the graphs of smoothed counts of PIP gaps first increase, then attain a maximum at an abscissa which grows with the count limit $x$, and only then start to decrease exponentially (this behavior can be observed in figure 1 of [5]).

Table 8 presents the record gaps (also known as maximal gaps [26]) that were observed up to $10^{15}$. Since a large gap between primes very likely corresponds to a large gap between PIPs (having the large prime gap between their indices), the first ten occurrences of each prime gap up to $4 \times 10^{18}$, obtained as a collateral result of the third author's extensive verification of the Goldbach conjecture [7], were used to locate large gaps between PIPs. This was done as follows:

1. given an index $i$ (the first prime of a large prime gap), an approximation $\hat{p}_i$ of $p_i$ was found by solving $|\hat{\pi}(\hat{p}_i) - i| < 10$, where $\hat{\pi}(x)$ is the Riemann's formula for $\pi(x)$, truncated to the first one million complex conjugate zeros on the critical line, and with lower order terms replaced by simpler asymptotic approximations;

2. using the algorithm described in [10], $\pi(\hat{p}_i)$ was computed (this was by far the most time consuming step);

3. using a segmented sieve and using $\hat{p}_i$ as a starting point, going backwards if necessary, $p_i$, which is by construction a PIP, was located;

4. since the gap between indices was known a priori, the next PIP was also located, and the difference between the two was computed.

The maximal gap candidates above $10^{15}$ that resulted from this effort are presented in Table 8; below $10^{15}$, the results of Table 8 were reproduced exactly. Based on the data from these two tables, it appears that $h/\log^3 q(h)$ is bounded.
Table 3: Record gaps between PIPs up to $10^{15}$

<table>
<thead>
<tr>
<th>$h$</th>
<th>$q(h)$</th>
<th>$h$</th>
<th>$q(h)$</th>
<th>$h$</th>
<th>$q(h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>1380</td>
<td>3733111</td>
<td>7524</td>
<td>49868272577</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>1500</td>
<td>5188297</td>
<td>7648</td>
<td>57757941919</td>
</tr>
<tr>
<td>14</td>
<td>17</td>
<td>1766</td>
<td>5336477</td>
<td>7884</td>
<td>60381716303</td>
</tr>
<tr>
<td>18</td>
<td>41</td>
<td>1784</td>
<td>7244099</td>
<td>8994</td>
<td>93033321509</td>
</tr>
<tr>
<td>26</td>
<td>83</td>
<td>1800</td>
<td>9026123</td>
<td>9208</td>
<td>104673577891</td>
</tr>
<tr>
<td>30</td>
<td>127</td>
<td>1852</td>
<td>12818959</td>
<td>9328</td>
<td>215587773169</td>
</tr>
<tr>
<td>36</td>
<td>241</td>
<td>1998</td>
<td>21330371</td>
<td>10254</td>
<td>271208089553</td>
</tr>
<tr>
<td>48</td>
<td>283</td>
<td>2200</td>
<td>21459517</td>
<td>10702</td>
<td>521584527307</td>
</tr>
<tr>
<td>92</td>
<td>617</td>
<td>2280</td>
<td>24931771</td>
<td>12388</td>
<td>655650146791</td>
</tr>
<tr>
<td>112</td>
<td>1297</td>
<td>2628</td>
<td>32637571</td>
<td>13436</td>
<td>1139727488171</td>
</tr>
<tr>
<td>114</td>
<td>1913</td>
<td>2992</td>
<td>79689091</td>
<td>13826</td>
<td>3565879112657</td>
</tr>
<tr>
<td>122</td>
<td>2099</td>
<td>3000</td>
<td>182395973</td>
<td>13898</td>
<td>514437850811</td>
</tr>
<tr>
<td>150</td>
<td>3761</td>
<td>3770</td>
<td>315619631</td>
<td>14570</td>
<td>8549998218191</td>
</tr>
<tr>
<td>168</td>
<td>5869</td>
<td>4406</td>
<td>390002363</td>
<td>15102</td>
<td>8724860034481</td>
</tr>
<tr>
<td>190</td>
<td>9103</td>
<td>4506</td>
<td>2199880757</td>
<td>15218</td>
<td>12118597117331</td>
</tr>
<tr>
<td>348</td>
<td>10909</td>
<td>4872</td>
<td>2515605941</td>
<td>16006</td>
<td>1347916888087</td>
</tr>
<tr>
<td>372</td>
<td>46751</td>
<td>4938</td>
<td>3443579963</td>
<td>16814</td>
<td>31885486594523</td>
</tr>
<tr>
<td>384</td>
<td>104827</td>
<td>5214</td>
<td>3994122787</td>
<td>17010</td>
<td>36971628663863</td>
</tr>
<tr>
<td>458</td>
<td>114089</td>
<td>5256</td>
<td>4043156627</td>
<td>18312</td>
<td>40798355884309</td>
</tr>
<tr>
<td>474</td>
<td>152953</td>
<td>5844</td>
<td>6111419117</td>
<td>19680</td>
<td>60481825851197</td>
</tr>
<tr>
<td>498</td>
<td>177791</td>
<td>5974</td>
<td>8440859467</td>
<td>21820</td>
<td>8104055147807</td>
</tr>
<tr>
<td>642</td>
<td>219931</td>
<td>6486</td>
<td>9568037147</td>
<td>22804</td>
<td>22992915352703</td>
</tr>
<tr>
<td>738</td>
<td>293123</td>
<td>6864</td>
<td>2147240259</td>
<td>24658</td>
<td>452388122520163</td>
</tr>
<tr>
<td>1028</td>
<td>368153</td>
<td>7098</td>
<td>29861568733</td>
<td>25694</td>
<td>647593721749763</td>
</tr>
<tr>
<td>1244</td>
<td>2101553</td>
<td>7446</td>
<td>35005449181</td>
<td>26148</td>
<td>804920613659501</td>
</tr>
</tbody>
</table>

Table 4: Potential record gaps between PIPs after $10^{15}$

<table>
<thead>
<tr>
<th>$h$</th>
<th>$q(h)$</th>
<th>$h$</th>
<th>$q(h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>27524</td>
<td>1451492253702853</td>
<td>50932</td>
<td>1797828789776991187</td>
</tr>
<tr>
<td>27462</td>
<td>3031506479624729</td>
<td>51282</td>
<td>336720144283080467</td>
</tr>
<tr>
<td>31184</td>
<td>3149270807374079</td>
<td>51496</td>
<td>5303163766511877793</td>
</tr>
<tr>
<td>33348</td>
<td>19759035095377103</td>
<td>54766</td>
<td>5948139313109849407</td>
</tr>
<tr>
<td>34428</td>
<td>19843157450987771</td>
<td>55438</td>
<td>8686480782592200319</td>
</tr>
<tr>
<td>34902</td>
<td>44370362884634417</td>
<td>56964</td>
<td>1313156851050612637</td>
</tr>
<tr>
<td>35560</td>
<td>48210577082615809</td>
<td>57744</td>
<td>1447137227453980343</td>
</tr>
<tr>
<td>35964</td>
<td>58458364312779077</td>
<td>60646</td>
<td>15209204300586561877</td>
</tr>
<tr>
<td>36225</td>
<td>63536060873650711</td>
<td>62244</td>
<td>181086189703357989</td>
</tr>
<tr>
<td>45390</td>
<td>63775464504542041</td>
<td>65278</td>
<td>3537628815644951609</td>
</tr>
<tr>
<td>46910</td>
<td>770359508644782761</td>
<td>67136</td>
<td>635263029820676603</td>
</tr>
<tr>
<td>46948</td>
<td>1186416917758809991</td>
<td>67236</td>
<td>146174033905511020897</td>
</tr>
<tr>
<td>47833</td>
<td>1263062213472998429</td>
<td>67356</td>
<td>170912819272488312527</td>
</tr>
</tbody>
</table>
9. A Goldbach-Like Conjecture for PIPs

Let \( R(n) \) be the number of pairs \((q, n - q)\) such that both \( q \) and \( n - q \) are PIPs. Just like for the classical Goldbach conjecture [27], the identity

\[
\sum_{n=1}^{L} R(n) x^n = \left( \sum_{q \leq L} x^q \right)^2 \mod x^{L+1},
\]
coupled with a fast polynomial multiplication algorithm based on the Fast Fourier Transform, makes it possible to compute \( R(n) \) for all \( n \leq L \) using only \( O(L^{1+\epsilon}) \) time and space. For \( L \) a positive even integer, let

\[
R_{\text{lower}}(x; L) = \min_{x \leq 2n \leq L} R(2n) \quad \text{and} \quad R_{\text{upper}}(x) = \max_{n \leq x} R(n).
\]

For \( n \) even, these two non-decreasing functions are useful lower and upper bounds of the value of \( R(n) \). Figure 9 shows how these two functions behave (their points of increase up to \( L = 10^9 \) were computed with the help of a simple matlab script). Based on our empirical data, the following conjecture is almost certainly true.

Conjecture 9.1. All even integers larger than 80612 can be expressed as the sum of two prime-indexed primes.
10. Proofs of Lemmas

Proof of Lemma 2.1. First, note that $n \log n = n^{1 + \log \log n}$. From this,
\[
\log (\log (n \log n)) = \log \left( \log n^{1 + \frac{\log \log n}{\log n}} \right) = \log \left( 1 + \frac{\log \log n}{\log n} \right) + \log \log n.
\]
Since $\log (1 + x) < x$ for $x > 0$, we have $\log \left( 1 + \frac{\log \log n}{\log n} \right) < \frac{\log \log n}{\log n}$, which completes the lemma’s proof.

Proof of Lemma 2.2. First, note that
\[
\log (n \log n) = n \log \left( n^{1 + \frac{\log \log n}{\log n}} \right) = (n \log n) \left( 1 + \frac{\log \log n}{\log n} \right).
\]
Now,
\[
1 + \frac{\log \log n}{\log n} = n \frac{\log \left( 1 + \frac{\log \log n}{\log n} \right)}{\log n} < n \frac{\log \log n}{\log n}
\]
because $\log (1 + x) < x$ for $x > 0$. Thus,
\[
\log (n \log (n \log n)) = \log \log \left( n \log n \left( 1 + \frac{\log \log n}{\log n} \right) \right)
< \log \log n^{1 + \frac{\log \log n}{\log n} + \log \log n} + \log \log n^{1 + \frac{\log \log n}{\log n} + \log \log n} = \log \log n + \log \left( 1 + \frac{\log \log n}{\log n} + \frac{\log \log n}{\log^2 n} \right)
< \log \log n + \frac{\log \log n}{\log n} + \frac{\log \log n}{\log^2 n},
\]
where the last inequality comes from again using $\log (1 + x) < x$. This proves the lemma.

Proof of Lemma 2.3. We begin with the upper bound. From [7], we know that $\pi(x) \geq \frac{x}{\log x}$ for all $x \geq 17$. Thus, in this range,
\[
\log \pi(x) \geq \left(1 - \frac{\log \log x}{\log x}\right) \log x,
\]
and so
\[
\frac{1}{\log \pi(x)} < \frac{1}{\log x} \left( 1 - \frac{1}{\log \log x} \right). \tag{10.1}
\]
Writing the second factor as a geometric series proves the bound.
Considering the lower bound, we use the upper bound on $\pi(x)$ for $x > 1$ in (4.2) to see that
\[
\log \pi(x) \leq \log \left( \frac{x}{\log x} \left( 1 + \frac{1.2762}{\log x} \right) \right) = \log \frac{x}{\log x} + \log \left( 1 + \frac{1.2762}{\log x} \right) \\
= \left( 1 - \frac{\log \log x}{\log x} \right) \log x + \log \left( 1 + \frac{1.2762}{\log x} \right) \\
< \left( 1 - \frac{\log \log x}{\log x} \right) \log x + 1.2762 \frac{\log x}{\log x},
\]
where the last inequality uses $\log(1 + x) < x$ for $x > 0$.

Taking the reciprocal of this inequality, we have
\[
\frac{1}{\log \pi(x)} \geq \frac{1}{\log x} \left( 1 - \frac{1}{\frac{\log \log x}{\log x} + 1.2762 \frac{\log x}{\log x}} \right).
\]

Thus, to establish the lemma, we need to bound
\[
\frac{1}{1 - \frac{\log \log x}{\log x} + 1.2762 \frac{\log x}{\log x}} \geq 1 + \frac{\log \log x}{\log x}. \tag{10.2}
\]

This is indeed the case, as we may rewrite the fraction as the sum of a geometric series. That is, we may write
\[
\frac{1}{1 - \frac{\log \log x}{\log x} + 1.2762 \frac{\log x}{\log x}} = 1 + \sum_{k=1}^{\infty} \left( \frac{\log \log x}{\log x} - \frac{1.2762}{\log x} \right)^k.
\]

Now, for $x \geq 33$,
\[
\sum_{k=2}^{\infty} \left( \frac{\log \log x}{\log x} - \frac{1.2762}{\log x} \right)^2 < \frac{1}{1 - \frac{\log \log x}{\log x} + 1.2762 \frac{\log x}{\log x}} > \frac{1.2762}{\log x},
\]

establishing (10.2). This completes the lemma’s proof.

**Proof of Lemma 2.4.** Using Lemma 2.3, together with (4.1), we have
\[
\frac{1}{\log^k \pi(x)} = \left( \frac{1}{\log x} \left( 1 + \frac{\log \log x}{\log x} + O \left( \frac{(\log \log x)^2}{\log^2 x} \right) \right) \right)^k \\
= \frac{1}{\log^k x} \left( 1 + \frac{k \log \log x}{\log x} + O_k \left( \frac{(\log \log x)^2}{\log^2 x} \right) \right), \tag{10.3}
\]
which establishes the first half of the lemma.

We know that
\[
\pi(x) = \frac{x}{\log x} \left( 1 + \frac{1}{\log x} + O \left( \frac{1}{\log^2 x} \right) \right),
\]
and so the same argument used in the proof of Lemma 2.2 gives that

\[
\log \log \pi(x) = \log \log x + \log \left( 1 - \frac{\log \log x}{\log x} + O \left( \frac{(\log \log x)^2}{\log^2 x} \right) \right)
\]

\[
= \log \log x + O \left( \frac{\log \log x}{\log x} \right).
\]

Using this and (10.3), we see

\[
\frac{\log \log \pi(x)}{\log \pi(x)} = \left( \log \log x + O \left( \frac{\log \log x}{\log x} \right) \right) \cdot \left( \frac{1}{\log x} \left( 1 + O \left( \frac{\log \log x}{\log x} \right) \right) \right)
\]

\[
= \frac{\log \log x}{\log x} + O \left( \frac{(\log \log x)^2}{\log^2 x} \right),
\]

completing the proof of the lemma. \qed

References